

STIC-Biotech/ChemLib

68491

Fr m: Li, Ruixiang
Sent: Monday, June 10, 2002 4:28 PM
To: STIC-Biotech/ChemLib
Subject: Sequence search of Application NO: 09/733,387

Please do a standard search on SEQ ID NOs: 43 and 44 against both the commercial and interference nucleic acid database.

Thank you very much!

Ruixiang Li
GAU 1646
CM1 10E18
Mail Box 10C01
306-0282

Edward Hart
Technical Info. Specialist
STIC/Biotech
CMI 6B02 Tel: 305-9203

RECEIVED
JUN 11 2002
STIC

	TYPE OF SEARCH:	VENDOR/COST(where applic.)
Searcher: _____	NA Sequences: _____	STN: _____
Phone: _____	AA Sequences: _____	DIALOG: _____
Location: _____	Structures: _____	Questel/Orbit: _____
Date Picked Up: <u>6/11/02</u>	Bibliographic: _____	DRLink: _____
Date Completed: <u>6/12/02</u>	Litigation: _____	Lexis/Nexis: _____
Searcher Prep/Review: _____	Full text: _____	Sequence Sys.: <u>A</u>
Clerical: _____	Patent Family: _____	WWW/Internet: _____
Online time: _____	Other: _____	Other (specify): _____

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 11, 2002, 20:11:27 ; Search time 1691.49 seconds
(without alignments)
13165.893 Million cell updates/sec

Title: US-09-733-387-43

Perfect score: 1650

Sequence: 1 atggcgagccagggcct.....actcgcgcatctcaagaatag 1650

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

- 1: em_estba:*
- 2: em_esthma:*
- 3: em_estin:*
- 4: em_estnu:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_hic:*
- 9: gb_estl:*
- 10: gb_est2:*
- 11: gb_hic:*
- 12: gb_gss:*
- 13: em_gss_hum:*
- 14: em_gss_inv:*
- 15: em_gss_pln:*
- 16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	613.4	37.2	779	10	BF183209 601809910
2	443.4	26.9	526	10	BM193875
3	439.4	26.6	736	10	EG975242 602843148
4	399	24.2	930	10	BF788639
c 5	394.2	23.9	436	10	BM145287
c 6	351.2	21.3	376	10	BM145497
7	338	20.5	585	10	BE281756
8	314	19.0	800	10	BI646880
9	309.2	18.7	518	9	AA981381
10	259.8	15.7	338	10	BE241639
c 11	229.6	13.9	391	10	U14114
12	216.4	13.1	1087	11	AK014445
c 13	207.8	12.6	432	10	BE831720
14	191.6	11.6	856	10	BI660803
15	168.4	10.2	329	10	BM484584
16	166.6	10.1	248	10	BM146119
17	151.8	9.2	3835	11	BC013207

18	145.8	8.8	1038	10	BM450054
19	143.6	8.7	953	10	BF164505
20	142.2	8.6	519	12	AQ171270
21	140.2	8.5	847	10	BI257601
22	140	8.5	567	12	AQ607309
23	139.8	8.5	816	10	BG760961
24	130	7.9	446	9	BB860353
25	119.4	7.2	823	10	BF345756
26	117	7.1	560	10	BI066761
27	116.2	7.0	677	10	BG818779
28	115.6	7.0	453	9	BB860092
29	114.2	6.9	537	10	BE754100
30	113.2	6.9	736	10	BI661416
31	111.4	6.8	431	9	AW592568
32	111.4	6.8	580	9	AI807052
33	110.2	6.7	790	10	BI691050
34	109.4	6.6	756	10	BI091842
35	108	6.5	600	10	BG804206
36	105.8	6.4	756	10	BG770361
37	105.8	6.4	934	10	BG032841
c 38	104.2	6.3	523	10	BG004188
39	104.2	6.3	883	10	BI755503
40	102.4	6.2	794	10	EG914136
41	99	6.0	659	10	BF783449
42	98.6	6.0	893	10	BF784683
43	97.8	5.9	753	10	BF584015
44	96.2	5.8	946	10	BF346638
45	95.4	5.8	733	10	BG833698

ALIGNMENTS

RESULT 1

BF183209 779 bp mRNA linear EST 31-OCT-2000
LOCUS 601809910F1 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:404042 5', mRNA sequence.
DEFINITION
ACCESSION BF183209.1 GI:11061396
VERSION
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 779)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgrabs-r@mail.nih.gov
Tissue Procurement: DCTD/DTP/Gazdar
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LUCM858 row: d column: 03
High quality sequence stop: 701.
Location/Qualifiers
1. 779
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:404042"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pOTB7; Site: 1: XhoI; Site 2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California,

FEATURES source

Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH-MGC Library."

```

BASE COUNT      136 a      229 c      218 g      196 t
ORIGIN

Query Match      37.2%; Score 613.4; DB 10; Length 779;
Best Local Similarity 97.5%; Pred. No. 2.4e-127;
Matches 676; Conservative 0; Mismatches 11; Indels 6; Gaps 5;

QY 662 ctaaaggaccactgagactgtctctgagggctgtccacgaggtgcacactgag 721
Db 2 CTAAGGGACCACTGGAGACTGGTCTGTGAGGGGTCTCCACGGAGGTGACAGCTGAGG 61
QY 722 ggaacctgtgtctgtgtgaccactgaccttttgcctctgtctgtgacccacttgg 781
Db 62 GGACCGTGTGTCTGTGTGACCACTGACCTTTTTCGCCCTGTCTCTGAGACCCACCTGG 121
QY 782 accagtccacggtgcatactccacacgcatctccagcggtgtgtgggtctccatga 841
Db 122 ACCATCCACGGTGCATATCTTCACACGCATCTCCACGGCGGGCTGTGGGFTCTCCATGA 181
QY 842 tctctgaccttaccattatttcttgcctttctgaggttttccggtgagaggttca 901
Db 182 TCTTCTGGCCTTACCACTATTCTTTATGCTTTCTGAGGCTTTCCCGGGAGAGGTTC 241
QY 902 agtcaagaatgcccccaagaatccacgtggtggtggtggtggtgtctctctgaatc 961
Db 242 AGTCAGAAGATGCCCAAGATCCACGTGG-CCTGGGTGGCAGCCTGTCTCTCTGAATC 300
QY 962 tggcctcttggtaagtggagtggtcctcaaaagggtctgagctgctgctgctggccc 1021
Db 301 TGGCCTTCTTGGTCAATGTGGGAGTGGCTCAAGGGGTCTGATGCTGCTGGGCC 360
QY 1022 ggggggtgtcttccactacttctgtctgtgcttccactggtggtggttgaagct 1081
Db 361 GGGGGGTGTCTTCCACTACTTCTGCTGTCTGTGCTTCCACTGGATGGGCTTGAAGCT 420
QY 1082 tccaccttaactgctgctgaggtcttcaacactaatttgggacactacttctctga 1141
Db 421 TCCACCTTACTCTGCTGCTGCTGAGGGTCTTCAACACCTACTTGGGCCTACTTCTCTGA 480
QY 1142 agctgagcgtgtgggtggggtggtggtggtggtggtggtggtggtggtggtggtggt 1201
Db 481 AGCTGAGCCTGTGGGTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCT 540
QY 1202 acagtaagcgtctacaccatccgtgatagggaacccgacactctctgagctatgct 1261
Db 541 ACAG--TACGGCTCTACACCATCGTGATAGGAGAAACGACCTCTCTGGAGCTATGCT 598
QY 1262 ggtt-cggtgaagggaacaacatgacgacctctatatccacgttccaggtacttctc 1320
Db 599 GGTTCCTGTGAAGGGGACAAACATGTAAGCGCTCTATATACCGCTCCACGG-TAATTCCTC 657
QY 1321 atcacttctcttggcatggtggtcctggcc 1353
Db 658 ATCA-CTTCTCTTTGGCATGGTGGTCTCTGGCC 689

RESULT 2
BM193875
LOCUS
DEFINITION
TCAAP1E4365 Pediatric pre-B cell acute lymphoblastic leukemia
Baylor-HGSC project-TCBA Homo sapiens cDNA clone TCAAP4365, mRNA
sequence.
BM193875
ACCESSION
VERSION
BM193875.1 GI:17652472
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 526)
 Wei,Y., Tsang,Y.T.M., Mei,G., Ku,J.M., Ali-Osman,F.R. Jr.,
 Gunaratne,P.H., Muzny,D., Bouck,J., Gibbs,R.A. and Margolin,J.F.
 Pediatric Leukemia cDNA Sequencing Project (2001)
 Unpublished (2001)
 Contact: Dr. Judith F. Margolin
 Texas Children's Cancer Center and Human Genome Sequencing Center
 at Baylor College of Medicine
 1102 Bates, MC3-3320 Houston, TX 77030, USA
 Tel: 832-824-4536
 Fax: 832-825-4038
 Email: clones@tccc.org
 Seq primer: M13 primer.
 Location/Qualifiers

```

1..526
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="TCAAP4365"
/clone_lib="Pediatric pre-B cell acute lymphoblastic
leukemia Baylor-HGSC project-TCBA"
/sex="male"
/tissue_type="leukopheresis"
/cell_type="pre-B cell"
/dev_stage="pediatric 2 years"
/lab_host="DH10B"
/note="vector: lambda pSB; Site_1: BamHI; Site_2: EcoRI;
First strand cDNA was primed with an anchored
XhoI-oligo(dT) primer [5'GGAGACTCGAGCGCGGAGGAG(T)VN
3': V=A,C,G; N=A,C,G,T] and then dg tailed. Second strand
was primed with a BamHI-dC primer
[5'AGAGAGCTGGATCGCGCGGCAATATATAT(C) 3'].
Double-stranded cDNA was then digested with BamHI and XhoI
and directionally cloned into the BamHI and Sali sites of
lambda pSB vector. Library went through one round of
normalization. Library was constructed by Wei Yu at RIKEN
of Japan (Carninci P., Westover A., Nishiyama Y., Ohsumi T.,
Itoh W., Nagaoka S., Sasaki Y., Muramatsu M.,
Schneider C., Hayashizaki Y., High efficiency selection of
full-length cDNA by improved biotinylated cap trapper.,
DNA Res 4: 1, 61-6, Feb 28, 1997)"
BASE COUNT      125 a      156 c      152 g      93 t
ORIGIN

```

```

Query Match      26.9%; Score 443.4; DB 10; Length 526;
Best Local Similarity 98.7%; Pred. No. 3.1e-89;
Matches 447; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 atggcgagccaggggctgtggggcctgtctctgtctctctctctctcaggt 60
Db 73 ATGGCGAGCCCGAGGGCTGTGGGGCTGTCTCTCTCTCTCTCTCTCTCTCTCT 132
QY 61 caggaaagccacgaaggccgaagaacacacctgctgggagcaacaacatgtacac 120
Db 133 CAGAAAGCCACCGAAGGCCCAAGAACACCTGCTCTGGGGAGCAACAATGTACGAC 192
QY 121 atcttcaactgaaatgacaaggctttgttcttcaacaaagtgcagcagtcgagcagc 180
Db 193 ATCTTCAACTTGAATGACAAGCTTTGTGCTTCCACCAAGTCAGGACGTCGGCAGCC 252
QY 181 tcttgcactgtgaaaaactgtcagagatactggtctaaactacgagggccatctgatgaag 240
Db 253 TCCCTGCAATGTGAAAACCTTCAGAGATACTTGGCTAAACTACGAGGCCCATCTGATGAAG 312
QY 241 gaaggtttgacgagaggtgaacacaccttctctgaaggtttgttgcagaacctcagc 300
Db 313 GAAGGTTTGACGCAAGAGGTGAACACGCTTCTCTGAAAGGCTTTGGTCCAGAACTCAGC 372
QY 301 accaactgtcagaagacttcttctctctgtgagccctctcaggttccgagggcagtg 360
Db 373 ACCAACACTGCAGAAGACTTCTATTCTCTCTGAGCCCTCTCAGGTTCCGAGCAGGTG 432
QY 361 atgaagacagaggaagcccccctgtacagagtgcacttcccaagagccttttctgatcc 420

```


BM145201/C

Db	249	GAGTGCCACACAGCTACGGCCTCTACACCATCGGTATAGGGAGAACCGCACCTCTCTGGGA	190
QY	1254	gctatgctgttccgtgaaggagacaacacatgaagccctctatatcacccgtccacggcta	1313
Db	189	GCTATGCTGTTCCGTGAGGGACAACCATGTACGCCCTCTATATACCGTCCACGGCTA	130
QY	1314	cttctcatcacctctctctcttggatggtggtctggtccct--ggtggtctggaagatc	1371
Db	129	CTTCTCATCATCTCTCTCTTGGCATGGTCTCTGAGACTAGAGTGGTCTGGAAGATC	70
QY	1372	ttcacctctccgtgctacacggtcaaggagcggggaagaaacccggaagaagtgctc	1431
Db	69	TTACCTCTCCGTGCTACAGCGTCAAGGAGCGGGGGAAGAACCCGGAAGAAGTGCTC	10
QY	1432	accctgctg 1440	
Db	9	ACCTGCTG 1	
RESULT	6		
LOCUS	BM145497/C		
DEFINITION	TCRAP1D6491 Pediatric acute myelogenous leukemia cell (FAB M1)		
ACCESSION	BM145497		
VERSION	BM145497.1	GI:17163567	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	Wei Y., Tsang Y.T.M., Mei G., Ku J.M., Ali-Osman F.R. Jr., Gunaratne P.H., Muzny D., Bouck J., Gibbs R.A. and Margolin J.F.		
TITLE	Pediatric leukemia cDNA Sequencing Project (2001)		
JOURNAL	Unpublished (2001)		
COMMENT	Contact: Dr. Judith F. Margolin Texas Children's Cancer Center and Human Genome Sequencing Center at Baylor College of Medicine 1102 Bates, MC3-3320 Houston, TX 77030, USA Tel: 832-824-4536 Fax: 832-825-4038 Email: clones@ccc.org Seq primer: M13 primer.		
FEATURES	Location/Qualifiers		
SOURCE	1. 376 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_lib="TCRAP6491" /clone_lib="Pediatric acute myelogenous leukemia cell (FAB M1) Baylor-HGSC project="CAA" /sex="male" /tissue_type="leukopheresis" /cell_type="myeloid cell" /dev_stage="pediatric 6 years" /lab_host="DH10B" /note="Vector: lambda pSB; Site 1: BamHI; Site 2: EcoRI; First strand cDNA was primed with an anchored XhoI-oligo(dT) primer [5'GGAGGACTCGAGGGCGCGAGGAGAG(T)YN 3'; V=A,C,G; N=A,C,G,T] and then dG tailed. Second strand was primed with a BamHI-dC primer [5'AGAGCTCGGATCGCGCGCGCAATAATAAT(C) 3']. Double-stranded cDNA was then digested with BamHI and XhoI and directionally cloned into the BamHI and SalI sites of lambda pSB vector. Library went through one round of normalization. Library was constructed by Wei Yu at RIKEN of Japan (Carninci P, Westover A, Nishiyama Y, Ohsumi T, Itoh M, Nagaoka S, Sasaki, Okazaki Y, Muramatsu M, Schneider C, Hayashizaki Y, High efficiency selection of full-length cDNA by improved biotinylated cap trapper., DNA Res 4: 1, 61-6, Feb 28, 1997)"		
BASE COUNT	82 a	100 c	118 g

```

/note="Organ: lung; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"
186 c 140 q 160 t
a

```

	BASE COUNT	ORIGIN
1	0	0
2	0	0
3	0	0
4	0	0
5	0	0
6	0	0
7	0	0
8	0	0
9	0	0
10	0	0
11	0	0
12	0	0
13	0	0
14	0	0
15	0	0
16	0	0
17	0	0
18	0	0
19	0	0
20	0	0
21	0	0
22	0	0
23	0	0
24	0	0
25	0	0
26	0	0
27	0	0
28	0	0
29	0	0
30	0	0
31	0	0
32	0	0
33	0	0
34	0	0
35	0	0
36	0	0
37	0	0
38	0	0
39	0	0
40	0	0
41	0	0
42	0	0
43	0	0
44	0	0
45	0	0
46	0	0
47	0	0
48	0	0
49	0	0
50	0	0
51	0	0
52	0	0
53	0	0
54	0	0
55	0	0
56	0	0
57	0	0
58	0	0
59	0	0
60	0	0
61	0	0
62	0	0
63	0	0
64	0	0
65	0	0
66	0	0
67	0	0
68	0	0
69	0	0
70	0	0
71	0	0
72	0	0
73	0	0
74	0	0
75	0	0
76	0	0
77	0	0
78	0	0
79	0	0
80	0	0
81	0	0
82	0	0
83	0	0
84	0	0
85	0	0
86	0	0
87	0	0
88	0	0
89	0	0
90	0	0
91	0	0
92	0	0
93	0	0
94	0	0
95	0	0
96	0	0
97	0	0
98	0	0
99	0	0
100	0	0

Query Match	20.5%	Score 338;	DB 10;	Length 585;
Best Local Similarity	76.1%;	Pred. No. 1.5e-05;		
Matches 446;	Conservative 0;	Mismatches 130;	Indels 10;	Gaps 2;
QY	591	ctcggagatcgtctctctcaccagcagcgcgcctcaacatgacctcacctcgtgtatt	650	
Db	1	TGTAGAGATCACTTCTCCCATGAGCGTCAACACCTAACATGATCCTCACCTGTGTATT	60	
QY	651	ctgggatgtgactaaaggagcaactgagagactggctctctgagggtgtctccacaggaagt	710	
Db	61	CTGGGATATGGCTAAA-----GGAGACTGGGATTCACAGGGTGTCTCCACGGTGCC	111	
QY	711	cagacctgagggaacgctgtgctgtgacacactgacctttttgcgcctgctcctgag	770	
Db	112	TGSGGATGGGAACTGCTGCCGTTCGACCACTTGACCTTCTTCGCCCTTGCTGCTGAG	171	
QY	771	accacacttgacactccagggtgcatactcaccgcgatcccaagcgaggctgtgg	830	
Db	172	GCCGATCTTGGACCTTGGCCAGGGACACAGACTCTCACCATCGCATCTCCACGAGGACGATGC	231	
QY	831	ggctccatgatcttccctggcctccaccattattctttatgcctttctgagcctttcccg	890	
Db	232	CGTCTCCATGATCTTCCCTGGCCTTCACCATGGTGCTCTATGTGGCTTCAGGTCTCTCT	291	
QY	891	ggagaggttcaagtacagaagatgcccaagatccacgtggccctgggtggcagcctgtt	950	
Db	292	GCAGAGGTTCAAGTCTGAAGATGCCCTAAGATCCACATGGCTCTGAGCATCAGCCTGTT	351	
QY	951	cctcctgaatctggcctctcttggctoaatgtggggagtggctcaaaagggtctgatgtgc	1010	
Db	352	TCTCTTGAATCTTACCTTCCCTGATCAATGTGGGAGCAGCTCCCAAGGCCCTCAGCCTC	411	
QY	1011	ctgcttggccgggggggtgcttctccactacttctcctctctgtgccttcaccttgatggg	1070	
Db	412	CTGCTGGGTCCGAGTGCATTTTCCACTACTTTTCTGCTGTGTCTTACCTTGGATGG	471	
QY	1071	ccttgaagccttccacctctaccctctgcctgctcagggtcttcaaacactactctgggca	1130	
Db	472	CCGTGAGGCGCTTCCACCTCTACCTGCTGGCATCAGAGCTTCAATCCCTACTCTTGGACA	531	
QY	1131	cta-cttctcgaagctgagcgtggtgggctggggcctgcgcgcct	1175	
Db	532	CTATTTTCTCTGAAGCTGAGCTGTGGCTGTGGCCTGGGCGCTGCCTGTCTCT	577	

RESULT 8

BI646880 800 bp mRNA linear EST 12-SEP-2001
603278628f1 NIH_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5318839 5',
mRNA sequence.

mRNA sequence.
BI646880
BI646880.1 GI:15561116

EST.
B1040000:1

EST.
house mouse.

house mouse.
Miss Macaulay

Mus musculus. M.

Eukaryota; M

Mammalia; Eu-

1 (bases 1

NIH-MGC <http>

National Ins

Unpublished

Computerized
Contact: Robert

Contact: robert@cs.cmu.edu
Email: cs.cmu.edu

Email: cgap@tiscali.it

Tissue Procurement

cDNA Library:

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
plate: LLAM11808 row: h column: 08
High quality sequence stop: 799.

FEATURES	SOURCE
----------	--------

```

location/Vuallillets
1. .800
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:5318839"
/clone_lib="NIH_CGAP_Mam3"
/tissue_type="tumor, gross tissue"
/lab_host="DH10B"
/notes="Organ: mammary; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert 2 kb. Library constructed by Life
Technologies, catalog #12017-018. Investigators providing
samples: Lothar Hennighausen/Chu-Xia Deng, NIH Reference
for transgenic model: Xu et al., Nature Genetics 22, 37-43
(1999). Note: this is a NCI_CGAP Library."
143 a 245 c 201 g 211 t
BASE COUNT
ORIGIN

```

Query Match 19.0%; Score 314; DB 10; Length 800;
Best Local Similarity 76.1%; Pred. No. 4.1e-60;
Matches 443; Conservative 0; Mismatches 125; Indels 14; Gaps 4;

QY	1072	cttgaagccttcacactctaccctgctcgtctgcagggtctctcaacactctctcggaac	1131
Db	1	CTGAGGCCCTCCACCTCTACCTGCTGGCCATCAGAGCTTCAATACCTACTT-GGACAC	59
QY	1132	tactctgaagctgaagctggctgggctgggctgcccctgatggctcatcggaact	1191
Db	60	TATTCTCTGAAGCTGAGCCTGCTGGCCTGGGGCTGCCTGTTCTGTAGTTAT-GGTGCT	118
QY	1192	ggaggtgccaacagctacggcctctacacacatccgtgataggagaaaccgcacctctctg	1251
Db	119	GGAAGTAGCAACATTTATGAGGTTTACACCATTCGGCACAGAGAAACCGCATCACTG	178
QY	1252	gagctatgctgggtccgtgaaggagacaacctgatcgcctctctatatacgcgtccacggc	1311
Db	179	GAGTGTGCTGGTCCAGAAAAGACC-----TGCTCTTTATGCTACTGTGCCACGGC	229
QY	1312	tactctcatcaacctctctctcttggcaataggtcctggccctggctggtctggaagatc	1371
Db	230	TACTTCCTGTATTACTTCCCTCTTCGGGNGCTGTGGTATGTGGCTTTGGGGCTTGAAGATC	289
QY	1372	ttcacacctgtcccgctctacagcgggtcaagagcgggggaagaacccgaagaaggtctc	1431
Db	290	TTCACTCTGCCCATGTCTACAGGGGGCAAGGACAGGGCCGACCTGCGAAGTCGGTCTC	349
QY	1432	acctgtcggcctctcagacctgtgtgggtgtacatgggggttgccactctcaacccg	1491
Db	350	ACCGTGTGGGGCTCTCCAGTCTGGTGGGATGACCTGGGGGGTGGCTGTCTCACCCCC	409
QY	1492	ttgggctctccacgctctacatctttgcactttcaactccttgcgaaggtgtcttcac	1551
Db	410	CTCGGCTGTGTACCATCTACGTCTTCAACCCTCTTAAACTCTCTACAAAGGCGCTTCTCATC	469
QY	1552	tgctgtggttcaccatcttttacctcccgaagtcagagaccacagtctctcctc--t	1608
Db	470	TTCTGCTGGTTTCATCTCTCTACTTCCCTACTCAGAGACACACAGGCTCTCTCTCCGGC	529
QY	1609	actgcaagattggaccaggccactccgcattctcaagaatag	1650
Db	530	ACGCCCCGCTGGACCAAGGCCACTCCGCTGTCCAGAGAGTAG	571

RESULT 9
AA981381

```

LOCUS      AA981381      518 bp      mRNA      linear      EST 27-MAY-1998
DEFINITION vx56d06.r1 Stratagene mouse macrophage (#937306) Mus musculus cDNA
            clone IMAGE:1279211 5' similar to TR:000406 000406 SEVEN
            TRANSMEMBRANE-DOMAIN RECEPTOR. ; mRNA sequence.
ACCESSION  AA981381
VERSION    AA981381.1 GI:3159917
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 518)
AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilison,R. and
            Waterston,R.
TITLE     The WashU-HHMI Mouse EST Project
JOURNAL   Unpublished (1996)
COMMENT   Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:671011
            Seq primer: -28ml3 rev1 ET from Amersham
            High quality sequence stop: 439.
FEATURES   Location/Qualifiers
            1..518
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /clone="IMAGE:1279211"
            /clone_lib="Stratagene mouse macrophage (#937306)"
            /tissue_type="macrophage"
            /dev_stage="WEHI-3 cell line"
            /lab_host="SOLR (kanamycin resistant)"
            /note="Organ: blood; Vector: pBluescript SK-; Site_1:
            EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
            Oligo dT. WEHI-3 cell line. Average insert size: 1.5 kb;
            Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGCAG
            3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"
BASE COUNT 85 a 152 c 131 g 150 t
ORIGIN
Query Match 18.7%; Score 309.2; DB 9; Length 518;
Best Local Similarity 75.9%; Pred. No. 4.1e-59;
Matches 399; Conservative 0; Mismatches 118; Indels 9; Gaps 1;

QY 1050 ctgtgcctcacctggatggccttggaagccttcacacctctacacctctgcgtgcagggt 1109
      |||||
Db 1 CTGTGCTTCCACCTGGATGGCTCTGGAGGCTTCCACCTCTACCTCTGCTGGCCATCAGAGT 60

QY 1110 cttaacacactctctggcactactctcctgaagctgagcctggtggcctgccc 1169
      |||||
Db 61 CTTCAATACCTACTTTTGGACACTATTTCTTGAAGCTGAGCTGCTGGCCTGGGCTGGCG 120

QY 1170 cgcctgaatgcatcgccactggagtgcaacagctcagcctctacacattccgtga 1229
      |||||
Db 121 TGTTCCTTGTAGTATTGGTCTGGAAGTAGCAGAGTTATGGAGTTTACACCAATCCGCA 180

QY 1230 tagggagaacgcacactctctgagctatgctggttcctggaagggaacacattgtacgc 1289
      |||||
Db 181 CCAGGAGAACCGACATCACTGGAGTTGTGCTGTTCCAGAAAGAGCC-----TGC 231

QY 1290 cctctatatcacgcctcagctactctctcctacacactcctctcttggcattggtgctc 1349
      |||||
Db 232 TCATTATGCTACTGTCCACGGCTACTTCCCTTGTACCTTCTCTCTCGTGTGCTGTTACT 291

```

```

QY 1350 ggccctgggtggtctggaagatcttccacctgtcccggtgtacagcggtcaaggagcggg 1409
      |||||
Db 292 GGTTTGGTGGCTTGGAGATCTTCACTCTGCCAGTGTTCACAGCGGCAAGACAGGG 351

QY 1410 gaagaacccggaagaaggtgctcaccctgtggcctcctgagcctggtgggtgtgacatg 1469
      |||||
Db 352 GCGGACCTGGAAGTTCGGCTCTCCCGTCTGGGCTCTCCAGTGTGGGCGATGACCTG 411

QY 1470 ggggtggccatcttcaacccgttggcctctccacgctctacatcttgcacatttcaa 1529
      |||||
Db 412 GGGGTGGCTGTCTCTACCCCTCGGCTTGTCTTACCATCTACGCTCTACCCCTCTTAAA 471

QY 1530 ctcttcgaaggtgtcttctatctgctgctggttcaccacctcttac 1575
      |||||
Db 472 CTCTCTACAAAGGCTCTTCTATCTTCTGCTGTTTCATCATCTCTTAC 517

RESULT 10
LOCUS      BE241639      338 bp      mRNA      linear      EST 03-OCT-2001
DEFINITION TCAAP1E0686 Pediatric acute myelogenous leukemia cell (FAB M1)
            Baylor-HGSC project-TCAA Homo sapiens cDNA clone TCAAP0686, mRNA
            sequence.
ACCESSION  BE241639
VERSION    BE241639.1 GI:9093362
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 338)
AUTHORS   Wei,Y., Tsang,Y.T.M., Mei,G., Ku,J.M., Ali-Osman Jr.,F.R., Muzny,D.
            Bouck,J., Gibbs,R.A. and Margolin,J.F.
            Pediatric Leukemia cDNA Sequencing Project
            Unpublished (2000)
            Contact: Dr. Judith F. Margolin
            Texas Children's Cancer Center and Human Genome Sequencing Center
            at Baylor College of Medicine
            1102 Bates, MC3-3320 Houston, TX 77030, USA
            Tel: 832-824-4536
            Fax: 832-825-4038
            Email: clones@ccc.org
            Citation: Carninci,P. and Hayashizaki,Y. High efficiency
            full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
            Seq primer: M13 primer.
FEATURES   Location/Qualifiers
            1..338
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="TCAAP0686"
            /clone_lib="Pediatric acute myelogenous leukemia cell (FAB
            M1) Baylor-HGSC project-TCAA"
            /sex="male"
            /tissue_type="leukopheresis"
            /cell_type="myeloid cell"
            /dev_stage="pediatric 6 years"
            /lab_host="DH10B"
            /note="Vector: lambda PSB; Site_1: BamHI; Site_2: EcoRI;
            First strand cDNA was primed with an anchored
            XhoI-oligo(dT) primer [5'GGAGACTGTCAGCGCCGAGGAGGAG(T)VN
            3'; V=A,C,G; N=A,C,G,T] and then dG tailed. Second strand
            was primed with a BamHI-dC primer
            [5'AGAGCTCGATCCGCGCGCAATAATAAT(C) 3'].
            Double-stranded cDNA was then digested with BamHI and XhoI
            and directionally cloned into the BamHI and SalI sites of
            lambda PSB vector. Library went through one round of
            normalization. Library was constructed by Wei Yu at RIKEN
            of Japan (Carninci P., Westover A, Nishiyama Y, Ohsumi T,
            Itoh M, Nagaoka S, Sasaki N, Okazaki Y, Muramatsu M,
            Schneider C, Hayashizaki Y, High efficiency selection of
            full-length cDNA by improved blotting cap trapper.,
            DNA Res 4: 1, 61-6, Feb 28, 1997)"
BASE COUNT 83 a 97 c 101 g 57 t

```


[illegible]

AUTHORS	Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldmann, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
TITLE	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE	20202863
COMMENT	Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-RC0-MT0059-2000-021-a08&t3=2000-06-20&tt=1) Seq primer: puc 18 forward High quality sequence start: 27 High quality sequence stop: 432. Location/Qualifiers 1. .432 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_lib="WT0059" /dev_stage="Adult" /note="Organ: marrow; Vector: puc18; Site_1: Smal; Site_2: Smal; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 / 716 - Ludwig Institute for Cancer Research) profiles
FEATURES	
source	

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 11, 2002, 20:16:47 ; Search time 2132.3 Seconds
(without alignments)
16193.212 Million cell updates/sec

Title: US-09-733-387-43
Perfect score: 1650
Sequence: 1 atggagacgcccaggagcct.....actcgcacatctcaagataa 1650

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl : *

```

1: gb_da:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fa:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
8						

1	1650	100.0	1650	6	AX167301	AX167301	Sequence
2	1543	93.5	1566	6	AX167281	AX167281	Sequence
3	1454.6	88.2	1527	6	AX167261	AX167261	Sequence
4	1441	87.3	1458	6	AX167291	AX167291	Sequence
5	1430.6	86.7	1515	6	AX167271	AX167271	Sequence
6	1161	70.4	1203	6	AX167305	AX167305	Sequence
7	1054	63.9	1119	6	AX167285	AX167285	Sequence
8	1020	61.8	1020	6	AX167303	AX167303	Sequence
9	965.6	58.5	1080	6	AX167265	AX167265	Sequence
10	952	57.7	1011	6	AX167295	AX167295	Sequence
11	941.6	57.1	1068	6	AX167275	AX167275	Sequence
12	913	55.3	936	6	AX167283	AX167283	Sequence
13	824.6	50.0	897	6	AX167263	AX167263	Sequence
14	811	49.2	828	6	AX167293	AX167293	Sequence
15	800.6	48.5	885	6	AX167273	AX167273	Sequence
16	772.6	46.8	825	6	AX167307	AX167307	Sequence
17	747.8	45.3	2127	10	AF249738	AF249738	Mus musculus
18	665.6	40.3	741	6	AX167287	AX167287	Sequence
19	628.4	38.1	678	6	AX167259	AX167259	Sequence
20	577.2	35.0	702	6	AX167267	AX167267	Sequence
21	563.6	34.2	633	6	AX167297	AX167297	Sequence
22	553.2	33.5	690	6	AX167277	AX167277	Sequence
23	490	29.7	4036	6	AX167311	AX167311	Sequence
24	394	23.9	612	6	AX167309	AX167309	Sequence
25	287.4	17.4	78267	2	AC022585	Homo sapi	
26	287.4	17.4	157765	2	AC018552	Homo sapi	
27	287	17.4	528	6	AX167289	AX167289	Sequence
28	266.6	16.2	603	6	AX147806	AX147806	Sequence
29	207.2	12.6	161229	2	AC106681	Rattus no	
30	198.6	12.0	489	6	AX167269	AX167269	Sequence
31	185	11.2	420	6	AX167299	AX167299	Sequence
32	174.6	10.6	477	6	AX167279	AX167279	Sequence
33	161	9.8	3360	6	AX342667	AX342667	Sequence
34	159.4	9.7	1425	6	AX108481	AX108481	Sequence
35	159.4	9.7	1587	6	AX108469	AX108469	Sequence
36	159.4	9.7	3393	6	AX167242	AX167242	Sequence
37	159.4	9.7	3711	9	BC008770	BC008770	Homo sapi
38	156.2	9.5	2082	6	AX029423	AX029423	Sequence
39	156.2	9.5	2822	9	AX298003	AX298003	Sequence
40	156.2	9.5	2822	9	AX106858	AX106858	Homo sapi
41	156.2	9.5	2834	6	AX029424	AX029424	Sequence
42	155	9.4	3686	6	AX061726	AX061726	Sequence
43	155	9.4	3845	6	E54092	E54092	Novel gene
44	155	9.4	3845	6	HSA011001	HSA011001	Homo sapi
45	153.4	9.3	3774	6	AX136371	AX136371	Sequence

ALIGNMENTS

RESULT	1
AXI67301	
LOCUS	
Sequence	43 from Patent WO0142287.
DEFINITION	
AXI67301	
ACCESSION	
VERSION	AXI67301.1 GI:14596745
KEYWORDS	
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 1650)

BASE COUNT	313 a	507 c	447 g	383 t
ORIGIN				

[illegible]

Qy	1021	cggggggctgtcttccaactacttccctgctgtgcttccctggatcgggcccttgaagcc	1080
Db	1021	CGGGGGCTGTCTCCACTACTTCTTCCGTCTGTGCTTCACTTCCGTGGAGCC	1080
Qy	1081	ttcacacttaactgtctgctgtctcagggtttcaaacactaacttcgggaactacttcctg	1140
Db	1081	TTCACTCTTACCTGTCTGCTGTCTCAGGGTCTTCAACACCTACTTCGGGCACTACTTCTCTG	1140
Qy	1141	aagctgagcctggctgggctggccgcgcctgatggttcactcgactcggagatgcc	1200
Db	1141	AAGCTGAGCCTGGTGGCTGGGGCTTCCCGCCTGATGGTCTATCGCACTGGAGTGCC	1200
Qy	1201	aacagctagggcctctacaccatccctgtagggagaaacgcacactctctcggagctatgc	1260
Db	1201	AACAGCTAGGGCCTCTACACCATCCGTGATAGGAGAAACCGACCTCTCTGGAGCTATGCT	1260
Qy	1261	tggttccgtgaagggaaccatgtacgcccctctatatcacgtccacgctcacttcctc	1320
Db	1261	TGTTCCGTGAAGGACACCATGTAGGCCCTCTATATCACCCTCACGGCTACTTCTCTC	1320
Qy	1321	ataccttctcttttggcatgggttcctggccctgggtgctggaagatcttccacctg	1380
Db	1321	ATCACCTTCTCTTTGGCATGGTGGTCTTGGCCCTGGTGTCTGGAAGATCTTCAACCTG	1380
Qy	1381	tccctgctacagcgtcaagagcgggggaagaaacggaaaggaaggctcaccctgctg	1440
Db	1381	TCCCTGTCTACAGCGTCAAGAGCGGGGGAAGAACCGGAAGAGTGTCTCACTCTGCTG	1440
Qy	1441	ggcctcagagcctgggggtgtgacatgggggttggccatcttcaccccgctggggcctc	1500
Db	1441	GGCCTCTCAGCGCTGGTGGGTGTGACATGGGGGTTGGCCATCTTCAACCCGTTGGGCTC	1500
Qy	1501	tcacgcgtacatcttgcactttcaactccttgcaaggtgtcttcactcgtcgtcgtg	1560
Db	1501	TCACCGCTTACATCTTTGCATCTTTCAACTCTCTTGCAAGGTGTCTTATCTGCTGCTGG	1560
Qy	1561	ttcacatcctttactctcccaagtcagagcacacagtcctctctctactcgaagatg	1620
Db	1561	TTTCAACATCTTTACTCTCCAAGTCAGAGCACACAGTCTCTCTCTTACTGCAAGATTG	1620
Qy	1621	gaccagggccactccgcactctcaagaatg	1650
Db	1621	GACCAGGGCCACTCCGCATCTCAAGAAATG	1650
RESULT	2		
AX167281			
LOCUS	AX167281	1566 bp	DNA
DEFINITION	Sequence 23 from Patent WO0142287.		
ACCESSION	AX167281		
VERSION	AX167281.1	GI:14596735	
KEYWORDS	human.		
SOURCE	human.		
ORGANISM	homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 1566)		
JOURNAL	Donoho, G., Scoville, J., Turner, C.A., Friedrich, G., Zambrowicz, B.		
FEATURES	and Sands, A.T.		
source	Novel human membrane proteins and polynucleotides encoding the same		
	Patent: WO 0142287-A 23 14-JUN-2001;		
	Lexicon Genetics Incorporated (US)		
	Location/Qualifiers		
	1..1566		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
BASE COUNT	294 a 475 c 439 g 358 t		
ORIGIN			

Query Match 93.5%; Score 1543; DB 6; Length 1566;
Best Local Similarity 100.0%; Pred. No. 9.4e-308;

[illegible][illegible]

Db	121	ATCTTCAACTTGAAATGCAAGGCTTTGTGCTTCAACCAAGTGCAGGCAGTCTCGGGCAGCGAC	180
Qy	181	tcctgcaatggaataactgcagagatactggtctaaactacagagcccaatctgatgaag	240
Db	181	TCTGTCAATGTGGAACACTTGCAGACATACTGGCTAACTACGAGGCCCATCTGATGAAG	240
Qy	241	gaaggttgacgcagaagtgaacacgccttccctgaagcctttggtccagaacctcaagc	300
Db	241	GAAGTTTGAACGAGAAAGTGAACAGCCCTTCTCTGAAGGCTTTGGTCCAGAACCTCAGC	300
Qy	301	accacactgcagaagactctattctctcctggagccctctcaggtttccgaggcaggtg	360
Db	301	ACCAACACTGCAGAAAGACTTCTATTTCCTCTCTGGAGCCCTCTCAGGTTCCGAGGCAGGTG	360
Qy	361	atgaagacgagacagaagccctcgaagatgcacitcccaagagccttttcgatcc	420
Db	361	ATGAGGACGAGGACAAAGCCCTGNACAGATGCACATTCCTCCAAAGGCCCTTTTCGATPCC	420
Qy	421	ctgcagggaacaggctgtgtgttcggttcggcgtcaacaattctggacattggtccaggg	480
Db	421	CTGCCAGGCAACAGCTGTGTGTCTCGGCTTGGCCGTGCACATTCCTGGACATTTGGTCCAGGG	480
Qy	481	actctctcaaggcccccgcgtcgcctcggagatggcagcgcggtgtgaacaactgc	540
Db	481	ACTCTCTTCAAGGCCCCCGCTGGGCTGGGAGATGCGAGGGCGCTGTTGAAACAATGCG	540
Qy	541	ctggtgggttgagtgggacaataatgaatgcatacgaactggctggagcctctggagatc	600
Db	541	CTGGTGGGTTTGAAGTGGGACAAATGCAATGCATGTCAACCAAGCTGGCTGGAGCCCTTGSAGATC	600
Qy	601	gtcttctcaccagcgacgcgccctcaactgaactcgaactctgtattctcggagatg	660
Db	601	GTCTTCTCTCACCAGCGACGCCGCCCTAACAATGACCCCTCACCTGTGTATTCGGGATGTG	660
Qy	661	actaaagggaacaactggagactgtgtctctgagggtctctcaacgaggtcagacctgaag	720
Db	661	ACTAAAGGGACCACTGGAGACTGGTCTTCTGAGGGCTGCTCCACGAGGTCAGACCTGAG	720
Qy	721	gggacggtgtcgtcgtgacacactgaactttctgcctcgtcctcgtcagaccacacttg	780
Db	721	GGGACGCGTGTGCTGTGACCACTGACCTTTTTCGCCCTGCTCCTCAGACCCACCTTG	780
Qy	781	gaccagtccaggtgatatactccacagcatctccacggcgggctgtggggtctcctag	840
Db	781	GACCATGCCAGGTGCATATCTCTCACACGATCTCCACGCGGGCTGTGGGGTCTTCATG	840
Qy	841	atcttctggccttcaccattattcttatgcctttctgaggtcttcccgggagaggttc	900
Db	841	ATCTTCTGTGCCCTTACCACTATTCTTTATGCGCTTCTTGAGGCTTTCCCGGGAGAGGTTC	900
Qy	901	aagtgcagaagatgccccaaagatccactggccctgggtgcaagcctgttctcctcgaat	960
Db	901	AAGTCAGAGATGCCCAAGATCCACGTGGCCCTTGGGTGGCAGGCTGTCTCTCTGAAT	960
Qy	961	ctggccttcttggtcaatgtggggagtgaggtcgaagggtctgatgtcctcgtcgtggcc	1020
Db	961	CTGGCCTTCTTGTGTCAATGTGGGGATGGCTCAAAAGGGGTGTGATGTGCTGTCTGGGCC	1020
Qy	1021	cggggggcgtctccactactctcgtcgtcgtccttcgaagggtctgatgtcctcgtgaagcc	1080
Db	1021	CGGGGGGCTGCTTCCACTACTTCTGTGCTGTGCTTCACTGGATGGGCTTGTGAAGCC	1080
Qy	1081	ttccacctctacctcgtcgtctgaagggtcttcaaacactactctgggcactactcctg	1140
Db	1081	TTCCACCTCTACCTGCTCGCTGTCAAGGTTCTTCAACACCTACTTTCGGGCACACTTCTCTG	1140
Qy	1141	aagctgagcctggtggcctgggcctgcgccctcgtatggatcgcgcactggagatgcc	1200
Db	1141	AAGCTGAGCCTGGTGGGCTTGGGGCTGCCGCCCTGATGTTGTCATCGGCACATGGGAGTGCC	1200
Qy	1201	aacagtacggcctctacaccatccgtgataggggagaaccgaacctctctggagctatgc	1260
Db	1201	AACAGTACGGCCTCTACACCATCGGTGATAGGAGAACCCAGCACTCTCTGGAGCTATGC	1260

Qy	1261	tggttccgtgaaggaggaacacatgaagccctctatatcacctgtccacggtccacgtactactctc	1320		
Db	1261	TGGTTCGGTGAAGGAGACAAACATGTACGCCCTCTATATCACCGTCCACGGCTACTTCTCT	1320		
Qy	1321	atcaccttccctcttggcatggtggtcgtccctggatggtctggaagatcttccacctg	1380		
Db	1321	ATCACCTTCCCTCTTTGGCATGSGTCTGTGCCCTGGTGTCTGGAAGATCTTCACCTGT	1380		
Qy	1381	tccgtgctacagcgggtcaagagcggggggaagaaacccggaagaaggtgctcacccctcgt	1440		
Db	1381	TCCCGTGTCTACAGCGGTCAAGAGCGGGGGAAGAACCGGAAGAGGTGCTCACCTGCTG	1440		
Qy	1441	ggcctctcgagcctggtgggt	1461		
Db	1441	GGCCTCTCGAGCCTTGCAAGT	1461		
RESULT 4					
AX167291	AX167291	1458 bp	DNA	linear	PAT 03-JUL-2001
LOCUS	Sequence 33 from Patent WO0142287.				
DEFINITION	AX167291				
ACCESSION	AX167291.1	GI:14596740			
VERSION					
KEYWORDS					
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	1 (bases 1 to 1458)				
	Donoho, G., Scoville, J., Turner, C.A., Friedrich, G., Zambrowicz, B.				
	and Sands, A.T.				
TITLE	Novel human membrane proteins and polynucleotides encoding the same				
JOURNAL	Patent: WO 0142287-A 33 14-JUN-2001;				
	Lexicon Genetics Incorporated (US)				
FEATURES	Location/Qualifiers				
source	1..1458				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
BASE COUNT	274 a	442 c	410 g	332 t	
ORIGIN					
Query Match	87.3%	Score 1441;	DB 6;	Length 1458;	
Best Local Similarity	99.5%;	Pred. No. 9.7e-287;			
Matches 1458;	Conservative	0;	Mismatches	0;	Indels 7; Gaps 1;
Qy	1	atggcgagcccgagggcctgaggccctactctctctctctctctctctctctctctcgcctcaggt	60		
Db	1	ATGGCGAGCCCGAGGGCCTTGGGGGCCCTGCTCTGCTCTCTCTCTCTCTCTCTCTCTCAGGT	60		
Qy	61	caggaaaagcccgaggggccaagaacacctgcctggggagcaacaacatgtacgac	120		
Db	61	CAGGAAAAGCCCGAGGAGGCCAAGAACACCTGCTCTGGGGAGCAACAACATGTACGAC	120		
Qy	121	atcttcaactgaatgacaaggcttctgtcttccacaaagtcgaggcagtcgagcagcag	180		
Db	121	ATCTTCAACTGATGACAGGCTTTGTCTTCCAAAGTGCAGGCAGTCTGGCAGCGAC	180		
Qy	181	tccgtcaatgtgaaaactgcagagatactggtctaaactacagagcccatctgatgaag	240		
Db	181	TCCGTCAATGTGAAAACCTTGCAGAGATCTGGCTAAACTACGAGGCCCATCTGATGAAG	240		
Qy	241	gaaggtttgacgcagaaggtgaacacgccttctctgaagccttgggtccagaacccctcagc	300		
Db	241	GAAGGTTTGAGCGAGAAGGTGAACACGCCCTTCTCTGAAGGCTTTGGTCCAGAACCTCAGC	300		
Qy	301	acaaactgcagaagactctatttctctctgagagccctctcaggttcgagagcgaggtg	360		
Db	301	ACCAACACTGCAGAAGACTTCTATTCTCTTGAGGCCCTCTCAGGTTCCGAGCGAGGTG	360		
Qy	361	atgaagcagcaggaacagccctcgacagagtcgcagcttcccaagagcctttttcgtacc	420		

Db 601 GTCCTTCTCACCAGCGCGCCCTAACATGACCCCTACCTGTGTATTCTGGGATGTG 660
Qy 661 actaaaggacactggagactggtctcttgagggtgctccacgaggtcagacactgag 720
Db 661 ACTAAGGAGCACCCTGGAGATGTCTCTGAGGGCTGTCTCCACGGAGGTACACCTGTAG 720
Qy 721 gggacggtgctgctgtgacacactgacatttttccctgctgctgacacacacttg 780
Db 721 GGGACCGTGTGTGTGTGACACCTGACCTTTTTCGCTGTCTGCTGTGACACCCACCTTG 780
Qy 781 gaccagtccaggtgcatatctcacgcatctccacgagcgcgagctggtggtctccatg 840
Db 781 GACCAGTCCAGGTGCATATCTTACACGATCTCCACGGGGCTGTGGGGTCTCCATG 840
Qy 841 atctcttggtccctcaccattattcttattgctgcttctgaggttttcccgaggaggttc 900
Db 841 ATCTTCTTGGCTTCCACATTAATTTTATGCTTCTGAGGCTTTTCCCGGGAGAGGTTC 900
Qy 901 asgtcagaagatgcccacaaagatccagtggtggtggtgagcctgttctccctgaat 960
Db 901 AGTTCAGAAATGCCCAAGATCCAGTGGCCCTGGGTGGCAGCCTGTCTCTCTGAAT 960
Qy 961 ctggcctctttgttcaatggtggagtggtctcaagggtgtctgatgctgctgctggcc 1020
Db 961 CTGGCCTTCTTGGTCAATGTGGGAGTGGCTCAAGGGGTCTGATGCTGCTGCTGGCC 1020
Qy 1021 cggggggctgtctccactacttctgctgtgcttccctcagctgagcttgagcc 1080
Db 1021 CGGGGGGTGTCTTCCACTACTTCTGCTCTGTGCTTCACTTGGATGGGCTTGAAGCC 1080
Qy 1081 ttccaccttaacctgctgctgctgaggtgttcaacacctacttccggcacttctctg 1140
Db 1081 TTCCACCTCTTACTCTGCTGTGAGGTCTTCAACACCTACTTGGGCACTACTTCTCTG 1140
Qy 1141 aagctgagctggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 1200
Db 1141 AGCTGAGCCTGTGTGGGTGGGGCTTCCCGCCCTGTATGCTATACCGTCCAGGCTACTTCCCTC 1200
Qy 1201 aacagctacggcctac 1260
Db 1201 AACAGCTACGGCTCTACACCATCGGTATAGGAGAACCGCACCTCTCTGGAGCTATGC 1260
Qy 1261 tggctccgtgagggagac 1320
Db 1261 TGSTTCCGTGAGGGAGAACACCATGACGCTCTATATACCGTCCAGGCTACTTCCCTC 1320
Qy 1321 atcaactctctcttggcatggtggtggtggtggtggtggtggtggtggtggtggtg 1380
Db 1321 ATCACTTCTCTTGGCATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1380
Qy 1381 tcccggtctacagcgttcaaggagcggggaagacacacacacacacacacacacacac 1440
Db 1381 TCCCGTGTACAGCGGTCAAGGAGCGGGGAAGAAC-----GCTGCTACCCCTGCTG 1433
Qy 1441 ggcctctcagcctggtggtggt 1461
Db 1434 GGCCTCTGAGCCTTGCAAGT 1454

RESULT 6
AX167305
LOCUS AX167305 1203 bp DNA linear PAT 03-JUL-2001
DEFINITION Sequence 47 from Patent WO0142287.
ACCESSION AX167305
VERSION AX167305.1 GI:14596747
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1203)
AUTHORS Donoho, G., Scoville, J., Turner, C.A., Friedrich, G., Zambrowicz, B.
and Sands, A.T.

TITLE Novel human membrane proteins and polynucleotides encoding the same
JOURNAL Patent: WO 0142287-A 47 14-JUN-2001;
Lexicon Genetics Incorporated (US)
FEATURES
source Location/Qualifiers
1..1203
/organism="Homo sapiens"
/db_xref="taxon:9606" 295 t
BASE COUNT 201 a 384 c 323 g 295 t
ORIGIN
Query Match 70.4%; Score 1161; DB 6; Length 1203;
Best Local Similarity 99.6%; Pred. No. 4.7e-229;
Matches 1161; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Qy 482 ctctctcaaggcggcccccgcgctgagagtgagcagcggcgtgttgaaacatgcc 541
Db 35 CCCCTTTCACAGGGCCCCCGCTCGGCTGGAGATGGCAGCGCGCTGTGAACAAATCGCC 94
Qy 542 tgggtggtttgagtggtggacaaatgcatgtcacaagctggtgagcctctggagatcg 601
Db 95 TGGTGGGTGTTGAGTGTGGGACAAATGATGTACCAAGCTGGCTGAGCCTCTGGAGATCG 154
Qy 602 tctctctcaccagcagcggcccccctaaacatgacccctcaccctgtattctgggagatga 661
Db 155 TCTTCTCTCACAGCAGCGCCCCCTAACATGACCTCACCCTGTGTATTCTGGGATGTGA 214
Qy 662 ctaaaaggacacactgagactggttctgagggctgtccacgaggtcagacactgagg 721
Db 215 CTAAGGGACCACTGGAGACTGGTCTTCTGAGGGCTGTCTCCACGGAGGTACAGCCTGAGG 274
Qy 722 ggaacgtgtgctgtgacacactgacctttttgcctctcctcgtcagacacacttgg 781
Db 275 GGACCGGTGCTGCTGTGACCACTGACCTTTTTCGCCCTGCTCTGAGACCCACCTTGG 334
Qy 782 accagtccacggtgcatatctcacaacacatctccagggcgtgtgaggtctccatga 841
Db 335 ACCAGTCCACGGTGTATATCTTCTACAGCATCTCCACGGCGGCTGTGGGTCTCCATGA 394
Qy 842 tctctctggtctcaccattattctttatgctttctgaggtttcccgaggagaggttca 901
Db 395 TCTTCTTGGCTTTCACCAATATTCTTTATGCTTTCTGAGGCTTTCGCGGGAGAGTTCA 454
Qy 902 agtcagaagatgcccacaaagatccacgtggtggtggtggtggtggttctctcctgaatc 961
Db 455 AGTCAGAGATGCCCAAGATCCAGTGGCCCTGGTGGGCAACGCTGTCTCTCTGAATC 514
Qy 962 tggcctcttgggtcaatggtggagtggtcacaagggtgtctgagctgctgctggtggccc 1021
Db 515 TGGCCTTCTTGGTCAATGTGGGAGTGGCTCAAGGGGTCTGATGCTGCTGCTGGGCC 574
Qy 1022 gggggggtctctccactacttctgctctgcttccactggtggtggtggtggtggtggt 1081
Db 575 GGGGGGTGTCTTCCACTACTTCTGCTGCTGCTTCACTTCCAGTGGGCTTGAAGCT 634
Qy 1082 tcaactctactgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1141
Db 635 TCCACCTCTTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 694
Qy 1142 agctgagcctggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 1201
Db 695 AGCTGAGCCTTGGTGGGCTGGGCCCTGCCGCCCTGATGCTGATGCTGCTGCTGCTGCTGCT 754
Qy 1202 acagctacggcctctac 1261
Db 755 ACAGCTACGGCCTCTACACCATCCCGGTATAGGAGAACCGCACCTCTCTGAGGATGCT 814
Qy 1262 ggttccgtgaaggac 1321
Db 815 GGTTCGCTGAAGGAGCAACCATGACGCCCTCTATATACCTCCTCCAGCTACTTCCCTCA 874
Qy 1322 tcaactctcttggcatggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 1381

Db	875	TCACCTTCTCTTTTGGCATGTTGGTCCCTGGCCCTGGTGGTCTGGAAGATCTTCACCCCTGT	934
QY	1382	cccggtctacagcgggtcaagggagcgggggaagaaacccggaagaagtgctcacccctgctgg	1441
Db	935	CCCGTGTCTACAGCGGTCAAGGAGCGGGGGAAGAACCGGGAAGAGTGCTCACCTGCTGG	994
QY	1442	gcctctgaagcctggtgggtgagatgggggttggccattctcaacccttgggacctct	1501
Db	995	GGCTCTCGAGCCTGGTGGGTGTGACATGGGGGTGGCCATCTTCACCCGCTGGGCGCTCT	1054
QY	1502	ccaccgtctacatcttgcacctttcaactcttgaactcttgaaggtgtctcatctgctgctgg	1561
Db	1055	CCACCGCTTACATCTTTGGACATTTTCACTCTTGCAAGGTGTCTTCATCTGCTGCTGGT	1114
QY	1562	tcaccatctttaccctcccaagtcagagacacacagtcctctctctactgcaagattgg	1621
Db	1115	TCACCATCTTTTACCTCCCAAGTCAGAGCACACAGTCTCCTCTCTACTTGAAGATTGG	1174
QY	1622	accaggcccaactccgcattctcaagaatag	1650
Db	1175	ACCAGGCCCACTCCGCATCTCAAGAAATAG	1203
RESULT 7			
LOCUS	AX167285	1119 bp	DNA
DEFINITION	Sequence 27 from Patent WO0142287.	linear	PAT 03-JUL-
ACCESSION	AX167285		
VERSION	AX167285.1		
KEYWORDS	GI:14596737		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostom		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 1119)		
JOURNAL	Doncho.G., Scoville,J., Turner,C.A., Friedrich,G., Zambrowicz,B.		
FEATURES	and Sands,A.T.		
Source	Novel human membrane proteins and polynucleotides encoding the		
BASE COUNT	Patent: WO 0142287-A 27 14-JUN-2001;		
ORIGIN	Lexicon Genetics Incorporated (US)		
	Location/Qualifiers		
	1..1119		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	182 a 352 c 315 g 270 t		
Query Match 63.9%; Score 1054; DB 6; Length 1119;			
Best Local Similarity 99.5%; Pred. NO. 5.2e-207;			
Matches 1057; Conservative 0; Mismatches 25; Indels 0; Gaps			
QY	482	ctctctcaaggcccccgctcgccctggagatggcagcggtgttgacaatgcc	541
Db	35	CCCTTTCACAGGCCCCCGCTCGGCTGGGAGATGGCAGCGGCTGTGAAACAAATGCC	94
QY	542	tggtgggtttgagtggtgggacaaatgcattgcacaaagctggctgagcctctggagatcg	601
Db	95	TGCTGGGTTTGAGTGTGGGACAAATGCATGTACCAAGCTGGCTGGAGATCG	154
QY	602	tctctctcaccagacgacccccccttaaatgacctcacctctgtattctgggatgta	661
Db	155	TCATTCTCACCAGCAGCCGCCCTAACATGACCTCACCTGTATTCTGGGATGTGA	214
QY	662	ctaaaggaccactgagagactggtctctgagggctgtctcaccgaggtcagacctgagg	721
Db	215	CTAAAGGGACCACTGGAGACTGGTCTTCTGAGGGGTGCTCCACGAGGTTCAGACCTGAGG	274
QY	722	ggaccgtgtgctgtgaccacctgaccttttgcgcctgctctctgagacccaccttg	781
Db	275	GGACCGTGTGCTGTGACACACTGACCTTTTTCGCCCTGCTCTCTGAGACCCACCTTGG	334
QY	782	accagtcaccggttgcatatctctcacacgcatctccagcgggggtgtggtggtctccatga	841

[illegible]

Qy	482	ctctcttcaaggagccccggctcgccctggagatgagcagcgctgtgtgaacaatcgcc	541
Db	35	CCCTTTACAGGCCCGCCGCTCGGCCTGGGAGATGGCAGCGCGCTGTGTAACAATCGCC	94
Qy	542	tggtgggttgagtgtaggacaaatgcattgtcaccaagctggctgagcctctggagatcg	601
Db	95	TGCTGGGTTTGTGTGGACAAATGCATGTACCAAGCTGGCTGAGCCTCTGGAGATCG	154
Qy	602	tcttctctcaacagagacagccccctaaacatgacctcaacctgtgtattcttgggattgaa	661
Db	155	TCATTCTCTACAGCAGCAGCCGCCCTTAACATGACCTCACTGTGTATTCTCGGATGTA	214
Qy	662	ctaaaggacacctgagactgagctgtctcttgagggtctgtctccacggaggttcagacctgagg	721
Db	215	CTAAGGGACCACTGGAGACTGGTCTTCTGAGSGCTGCTCCACGGAGGTCAGACCTGAGS	274
Qy	722	ggacctgtgctactgtgaccacctgaccttttttcgacctgtctctctgagacccaccttgg	781
Db	275	GGACCGGTGTCTGTGTGACCACTGACCTTTTTCGCCCTGTCTCTGTGACCACTCTGG	334
Qy	782	accagttcacgggtgcatactctcacacgcatctcccagggcgagctgtgggtctccatga	841

```

ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE     1 (bases 1 to 1020)
AUTHORS      Donoho,G., Scoville,J., Turner,C.A., Friedrich,G., Zambrowicz,B.
              and Sands,A.T.
TITLE         Novel human membrane proteins and polynucleotides encoding the same
JOURNAL       Patent: WO 0142287-A 45 14-JUN-2001;
              Lexicon Genetics Incorporated (US)
FEATURES      Location/Qualifiers
              source          1..1020
                           /organism="Homo sapiens"
                           /db_xref="taxon:9606"

```



```
Db 541 GCCTGATGGTCATCGGCACCTGGGAGTGCCAAAGCTACGGCTCTACACCATCGGTGAT 600
QY 1231 agggagaacgcacctctctgagctatgctgttccgtgaaggacacacattacgcc 1290
Db 601 AGGAGAACCGACCTCTCTGAGCTATGCTGTTCCGTCGAGGACAAACCAATGACGCC 660
QY 1291 ctctatacacgcgtccacggctactctccatcacctctctctctctctctctctctg 1350
Db 661 CTCTATATACCGCTCACCGCTACTTCCTCATACCTTCTCTTTGGCATGGTGTCTG 720
QY 1351 gccctgggtggaagattctcaacctgtcccgctgtacagcggtcaaggagcggggg 1410
Db 721 GCCTGGTGGTTCGGAAGATCTACCCCTGTCGCGTCTACAGCGGTCAAGGAGCGGGG 780
QY 1411 aagaacggaagaagtgctcacctgtggcctctcagcctcgtggtggtgacatgg 1470
Db 781 AAGAACCGNAGAGGTGTCACCCCTGTCGGGCTCTCGAGCCTGGTGGGTGTGACATGG 840
QY 1471 ggggtggccattctcaccccggttggcctctccacggtctacattttgacatttcaac 1530
Db 841 GGGTTGGCCATCTCACCCCGTGGGCCCTCTCCACCGTCTACATCTTTGCACTTTTCAAC 900
QY 1531 tccttgcaagggtg 1543
Db 901 TCCTTGAAGGTG 913

RESULT 13
AX167263
LOCUS AX167263 897 bp DNA linear PAT 03-JUL-2001
DEFINITION Sequence 5 from Patent WO0142287.
ACCESSION AX167263
VERSION AX167263.1 GI:14596726
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 897)
Donoho,G., Scoville,J., Turner,C.A., Friedrich,G., Zambrowicz,B.
and Sands,A.T.
Novel human membrane proteins and polynucleotides encoding the same
Patent: WO 0142287-A 5 14-JUN-2001;
Lexicon Genetics Incorporated (US)
FEATURES
Location/Qualifiers
Source
1..897
/organism="Homo sapiens"
BASE COUNT 154 a 273 c 250 g 220 t
ORIGIN
1
Query Match 50.0%; Score 824.6; DB 6; Length 897;
Best Local Similarity 99.5%; Pred. No. 9.6e-160;
Matches 827; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 631 atgacctcacctgttattctggatgtgactaaaggacacactgagactggtcttct 690
Db 1 ATGACCTCACCTGTGTATCTGGGATGTGACTAAAGGACCACTGGAGACTGGTCTTCT 60
QY 691 gagggtgtccacggaggtcagacctgaggggaccgtgtgctgtgacctgacc 750
Db 61 GAGGGTGTCTCCACGGAGGTGACACCTGAGGGGACCGTGTGCTGTGACCACTGACC 120
QY 751 ttttggccctgctcctgagaccacacttggaccagtccacgggtgcatactctcaacgc 810
Db 121 TTTTTCGCCCTGTCTGTGAGACCACTTGGACCACTGACCGTGCATATCTCTCACACGC 180
QY 811 atctccacggcggtgtgggtgtcctcatgatcttctggtccctccaccatttcttat 870
Db 181 ATCTCCAGCGGGGTGTGGGGTCTCCATGATCTTCTTGGCCCTTCACCAATTATCTTAT 240
QY 871 gcctttgtgaggttttcccggggagaggttcaadtcagaagatgcccccaagatccacgtg 930
```

```
Db 241 GCCTTCTGAGGCTTTCCTCGGAGAGGTTCAAGTCAGAAGATGCCCAAGATCCACGTG 300
QY 931 gccctgggtgacacgttctcctgaactgaactgaccttcttgggtcaatgtggagatggc 990
Db 301 GCCCTGGGTGGCAGCGTGTCTCTGAATCTGGCCCTTCCTGGTCAATGTGGGAGTGGC 360
QY 991 tcaaagggtgtgatgctgctgtggcccgggggggtgtcttccactacttctctgctc 1050
Db 361 TCAAGGGGTCTGATGCTGCTGCTGGCCCGGGGGGCTGTCTTCCACTACTTCTCTGCTC 420
QY 1051 tgtgcttcaacctggatggccttgaagccttccacctctacctgtctgtcgtcagggctc 1110
Db 421 TGTGCTTTCACCTGGATGGGCTTTGAAGCCTTCCACCTCTACCTGCTCGCTGTCAAGGTC 480
QY 1111 ttcaaacactacttcgggcactacttctgaagctgaagctgaagctgggtgggctgctccc 1170
Db 481 TTCAACACCTACTTCGGGCACTACTTCTGAAGCTGAGCCTGGTGGGCTGGGCTGCC 540
QY 1171 gccctgattgctcgcactgggagtgccacagctlacggcctctacacatccgtgat 1230
Db 541 GCCCTGATGGTTCATCGGCACCTGGGAGTGCCAAACAGCTACGGCCTTACACCATCCGTGAT 600
QY 1231 agggagaacgcacctctctgtgagctatgctgttccogtgaagggaacacatgacgcc 1290
Db 601 AGGGAGAACCACCTCTCTGAGCTATGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660
QY 1291 ctctatatccggtccacggctacttctcctcactcctctctctctctctctctctgctg 1350
Db 661 CTCTATATACCGTCCACGGCTACTTCTCATCACCTTCTCTTGGCATGGTGGTCTGCTG 720
QY 1351 gccctgggtgtggaagattctcacctgtcccgtgtctacagcggtcaaggagcggggg 1410
Db 721 GCCCTGGTGGTCTGGAAGATCTTACCCCTGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780
QY 1411 aagaacgggaaggtgtgtcaccctgtggcctctcgtggcctctcgtgctggtgggt 1461
Db 781 AAGAACCAGGAAGAGGTGCTCACCGTGTGGGGCTCTCGAGCCTTGCAAGT 831

RESULT 14
AX167293
LOCUS AX167293 828 bp DNA linear PAT 03-JUL-2001
DEFINITION Sequence 35 from Patent WO0142287.
ACCESSION AX167293
VERSION AX167293.1 GI:14596741
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 828)
Donoho,G., Scoville,J., Turner,C.A., Friedrich,G., Zambrowicz,B.
and Sands,A.T.
Novel human membrane proteins and polynucleotides encoding the same
Patent: WO 0142287-A 35 14-JUN-2001;
Lexicon Genetics Incorporated (US)
FEATURES
Location/Qualifiers
Source
1..828
/organism="Homo sapiens"
BASE COUNT 133 a 256 c 234 g 205 t
ORIGIN
1
Query Match 49.2%; Score 811; DB 6; Length 828;
Best Local Similarity 99.2%; Pred. No. 6.1e-157;
Matches 828; Conservative 0; Mismatches 0; Indels 7; Gaps 1;
QY 631 atgacctcacctgttattctggatgtgactaaaggacacactgagactggtcttct 690
Db 1 ATGACCTCACCTGTGTATCTGGGATGTGACTAAAGGACCACTGGAGACTGGTCTTCT 60
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 11, 2002, 20:20:01 ; Search time 212.41 Seconds
(without alignments)
13336.984 Million cell updates/sec

Title: US-09-733-387-43

Perfect score: 1650

Sequence: 1 atggcgacccagggcct.....actccgcatctcaagaatag 1650

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_032802.*

- 1: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
- 2: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
- 3: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
- 4: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
- 5: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
- 6: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
- 7: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
- 8: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
- 9: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
- 10: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
- 11: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
- 12: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
- 13: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
- 14: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
- 15: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
- 16: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
- 17: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
- 18: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
- 19: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
- 20: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
- 21: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
- 22: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
- 23: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
- 24: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Query Match %	Length	ID	Description
1	1650	100.0	1650	22 AAD08779	Human novel G-prot
2	1543	93.5	1566	22 AAD08769	Human novel G-prot
3	1454.6	88.2	1527	22 AAD08759	Human novel G-prot
4	1441	87.3	1458	22 AAD08774	Human novel G-prot
5	1430.6	86.7	1515	22 AAD08764	Human novel G-prot
6	1161	70.4	1203	22 AAD08781	Human novel G-prot
7	1054	63.9	1119	22 AAD08771	Human novel G-prot
8	1020	61.8	1020	22 AAD08780	Human novel G-prot
9	965.6	58.5	1080	22 AAD08761	Human novel G-prot

10	952	57.7	1011	22 AAD08776	Human novel G-prot
11	941.6	57.1	1068	22 AAD08766	Human novel G-prot
12	913	55.3	936	22 AAD08770	Human novel G-prot
13	824.6	50.0	897	22 AAD08760	Human novel G-prot
14	811	49.2	828	22 AAD08775	Human novel G-prot
15	800.6	48.5	885	22 AAD08765	Human novel G-prot
16	772.6	46.8	825	22 AAD08782	Human novel G-prot
17	665.6	40.3	741	22 AAD08772	Human novel G-prot
18	663.4	40.2	2433	23 AAS79005	DNA encoding novel
19	628.4	38.1	678	22 AAD08758	Human novel G-prot
20	611.6	37.1	811	22 ABA09625	Human bone marrow
21	611.4	37.1	1755	22 ABA09709	Human bone marrow
22	609.4	36.9	611	22 AAH98979	Human EST-derived
23	577.2	35.0	702	22 AAD08762	Human novel G-prot
24	563.6	34.2	633	22 AAD08777	Human novel G-prot
25	553.2	33.5	690	22 AAD08767	Human novel G-prot
26	495.2	30.0	2495	21 AAA78394	Human secreted pro
27	495.2	30.0	2502	22 AAD07673	Human secreted pro
28	490	29.7	4036	22 AAD08784	Human novel G-prot
29	472	28.6	1431	22 AAD07657	Human secreted pro
30	394	23.9	612	22 AAD08783	Human novel G-prot
31	380.2	23.0	838	22 AAD07674	Human secreted pro
32	287	17.4	528	22 AAD08773	Human novel G-prot
33	287	17.4	584	22 AAK37405	Human bone marrow
34	287	17.4	4992	22 AAK73522	Human immune/haema
35	287	17.4	4992	22 AAK73804	Human immune/haema
36	284	17.2	284	22 AAK50284	Human bone marrow
37	266.6	16.2	603	22 AAH50394	Human NGPCR38 codi
38	255	15.5	536	22 AAK58848	Human immune/haema
39	255	15.5	536	22 AAI63886	Human polynucleoti
40	231.4	14.0	1193	24 ABL01575	Human secreted pro
41	198.6	12.0	489	22 AAD08763	Human novel G-prot
42	195.6	11.9	3861	22 AAK73521	Human immune/haema
43	195.6	11.9	3861	22 AAK73805	Human immune/haema
44	185	11.2	420	22 AAD08778	Human novel G-prot
45	174.6	10.6	477	22 AAD08768	Human novel G-prot

ALIGNMENTS

RESULT 1

AAD08779

ID AAD08779 standard; cdna; 1650 BP.

XX AC AAD08779;

XX DT 04-SEP-2001 (first entry)

XX DE Human novel G-protein coupled receptor (NGPCR) #22.

XX KW Human; novel G-protein coupled receptor; NGPCR; signal transduction; drug screening; transmembrane protein; mental disorder; gene therapy; nontropic; pharmacogenomics; behavioural disorder; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX CDS 1..1650

XX FT /*tag= a

XX FT /product= "Human NGPCR protein"

XX PN WO200142287-A2.

XX PD 14-JUN-2001.

XX PF 07-DEC-2000; 2000WO-US33241.

XX PR 07-DEC-1999; 99US-0169427.

XX PA (LEXI-) LEXICON GENETICS INC.

XX PI Donoho G, Scoville J, Turner CA, Friedrich G, Zambrowicz B;

PI Sands AT;

XX WPI; 2001-381634/40.

DR P-PSDB; AAE04469.

XX

PT Novel polynucleotides encoding human G protein coupled receptors useful

PT for drug screening, diagnosis and in gene therapy of physiological or

PT behavioral disorders -

XX

PS Claim 1; Page 84; 91pp; English.

XX

CC The present sequence is a cDNA encoding human novel G protein coupled

CC receptor (NGPCR). NGPCRs are transmembrane proteins that span the

CC cellular membrane and are involved in signal transduction after ligand

CC binding. NGPCR DNA sequences are useful as hybridisation probes for

CC screening libraries and assessing gene expression patterns. Sequences

CC derived from regions adjacent to the intron/exon boundaries of NGPCR

CC gene can be used to design primers for use in amplification assays to

CC detect mutations within the exons, splice sites, introns that can be

CC used in diagnostics and pharmacogenomics. NGPCR sequences and/or gene

CC products are useful for identifying compounds that modulate gene

CC expression or gene product activity. Such compounds are useful in the

CC treatment of mental, physiological or behavioural disorders and diseases.

CC NGPCR DNA sequences and antibodies are useful for diagnostic and

CC prognostic evaluation of disorders related to NGPCR function and for the

CC identification of subjects having a predisposition to such disorders.

CC NGPCR DNA sequences are also useful for drug screening and in gene

CC therapy for modulating NGPCR expression and to produce genetically

CC engineered host cells to express NGPCR products in vivo. The encoded

CC NGPCR proteins are useful for generating antibodies, as reagents in

CC -diagnostic assays and for identifying other cellular gene products

CC related to NGPCR.

XX

SQ Sequence 1650 BP; 313 A; 507 C; 447 G; 383 T; 0 other;

Query Match 100.0%; Score 1650; DB 22; Length 1650;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1650; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atgagcagccagggcctggggcctgtctctgtctctctgtctctctgtctctcaggt 60

Db 1 atgagcagccagggcctggggcctgtctctgtctctctgtctctcaggt 60

Qy 61 cagaaagccaccagggcccaagaaacacgtgctggggagcaacaatgtacgac 120

Db 61 cagaaagccaccagggcccaagaaacacgtgctggggagcaacaatgtacgac 120

Qy 121 atctcaactggaatgacaaggctttgtcttaccagtgcaggaagtcggcagcgac 180

Db 121 atctcaactggaatgacaaggctttgtcttaccagtgcaggaagtcggcagcgac 180

Qy 181 tcttgcgaatgtggaacttcagagatactggttaactacagggccctctgtgaag 240

Db 181 tcttgcgaatgtggaacttcagagatactggttaactacagggccctctgtgaag 240

Qy 241 gaaggtttgacgagaagggaacacgcttctgtgaagctttgttccagaacctcagc 300

Db 241 gaaggtttgacgagaagggaacacgcttctgtgaagctttgttccagaacctcagc 300

Qy 301 accaactgacagaagacttctatttctctgtgagccctctcaggttccgagggcaggtg 360

Db 301 accaactgacagaagacttctatttctctgtgagccctctcaggttccgagggcaggtg 360

Qy 361 atgaagacagggacagcccttcagagatgogacttcccaagagccttttcgatcc 420

Db 361 atgaagacagggacagcccttcagagatgogacttcccaagagccttttcgatcc 420

Qy 421 ctgcagggcaacaggtctgtgttcgctgtgcgttcacacttctgacattgttcacagg 480

Db 421 ctgcagggcaacaggtctgtgttcgctgtgcgttcacacttctgacattgttcacagg 480

Qy 481 actctcttcaaggccccccgctcgctggggagatggcagcggtgtgtgaaacaatgcg 540

Db 481 actctcttcaaggccccccgctcgctggggagatggcagcggtgtgtgaaacaatgcg 540

Qy 541 ctggtgggtttgagtgaggacaaatgcatgtcaccaaagctggtgagcctctggagatc 600

Db 541 ctggtgggtttgagtgaggacaaatgcatgtcaccaaagctggtgagcctctggagatc 600

Qy 601 gtcttctcaccagcgacggccctcctaatacatgacctcacctgtgtattctggatgtg 660

Db 601 gtcttctcaccagcgacggccctcctaatacatgacctcacctgtgtattctggatgtg 660

Qy 661 actaaaggaccactgagactggtcttctgagggctgctccacggaggtcagacctgag 720

Db 661 actaaaggaccactgagactggtcttctgagggctgctccacggaggtcagacctgag 720

Qy 721 gggaccgtgtgctgtgaccacactgaccttttgcgcctgctctctgagaccaccttg 780

Db 721 gggaccgtgtgctgtgaccacactgaccttttgcgcctgctctctgagaccaccttg 780

Qy 781 gaccagtcacaggtgcatatctcacaacgcatctccagggcggtgtggtgtctccatg 840

Db 781 gaccagtcacaggtgcatatctcacaacgcatctccagggcggtgtggtgtctccatg 840

Qy 841 atcttctgacctcaccattatttctgcttctgagcttcccgagagaggttc 900

Db 841 atcttctgacctcaccattatttctgcttctgagcttcccgagagaggttc 900

Qy 901 agtgcagaagatgccccaagatccacgtggcctgggtggcagcctgttctctctgaat 960

Db 901 agtgcagaagatgccccaagatccacgtggcctgggtggcagcctgttctctctgaat 960

Qy 961 ctggccttcttggtaaatgtggggagtgctcaaaaggggtgtgatgtgctgtgggccc 1020

Db 961 ctggccttcttggtaaatgtggggagtgctcaaaaggggtgtgatgtgctgtgggccc 1020

Qy 1021 cggggggctgtcttccactactctctgtctgtcttccactgagtgagggccttgaagcc 1080

Db 1021 cggggggctgtcttccactactctctgtctgtcttccactgagtgagggccttgaagcc 1080

Qy 1081 ttccacctactctgctgctgctcaggggttcttcaacactactctcgggcaactactctg 1140

Db 1081 ttccacctactctgctgctgctcaggggttcttcaacactactctcgggcaactactctg 1140

Qy 1141 asgtgagcctggtgggtggggcctgcgcctcagctgagtgatcggcaactggagtgcc 1200

Db 1141 asgtgagcctggtgggtggggcctgcgcctcagctgagtgatcggcaactggagtgcc 1200

Qy 1201 aacagctacggcctctacacctcctgtatagggaacccgacactctctggagctatgc 1260

Db 1201 aacagctacggcctctacacctcctgtatagggaacccgacactctctggagctatgc 1260

Qy 1261 tgggttcgtgaaaggacaaacctgtacgcccctctataccacgttccacggtactctctc 1320

Db 1261 tgggttcgtgaaaggacaaacctgtacgcccctctataccacgttccacggtactctctc 1320

Qy 1321 atcaccttctcttggcagatggtgctgctggcctgggtgggtggaagatcttccacctg 1380

Db 1321 atcaccttctcttggcagatggtgctgctggcctgggtgggtggaagatcttccacctg 1380

Qy 1381 tccctgtctacagcggtcaagagcgggggaagaaacccggaagaaggtgctcacacctgctg 1440

Db 1381 tccctgtctacagcggtcaagagcgggggaagaaacccggaagaaggtgctcacacctgctg 1440

Qy 1441 ggcctctcagcctggtggtgtgacatgggggttggccattcttccacccgttggggcctc 1500

Db 1441 ggcctctcagcctggtggtgtgacatgggggttggccattcttccacccgttggggcctc 1500

Qy 1501 tccacgctacatcttgcaccttttaactccttgcagaggtgttcttatctgctgctg 1560

Db 1501 tccacgctacatcttgcaccttttaactccttgcagaggtgttcttatctgctgctg 1560

Qy 1561 ttccacactcttctacctcccaagtcagagcacacagctctctctctctacgtcagattg 1620

Db 1561 ttccacactcttctacctcccaagtcagagcacacagctctctctctctacgtcagattg 1620

QY 541 ctggtgggttgagtgaggacaaatgcatgacacaaagctgagcctctgagatc 600
|||||
Db 541 ctggtgggttgagtgaggacaaatgcatgacacaaagctgagcctctgagatc 600
QY 601 gtcttctcaccagcgagcccccctaaacatgacctcacctctgtattctggatgtg 660
|||||
Db 601 gtcttctcaccagcgagcccccctaaacatgacctcacctctgtattctggatgtg 660
QY 661 actaaaggacactgagagactgtcttctgaggggtgtctcacaggaggtcagacctgag 720
Db 661 actaaaggacactgagagactgtcttctgaggggtgtctcacaggaggtcagacctgag 720
QY 721 gggaccgtgtgtctgtgacacactgaccttttttcgacctctctctgagacccacctg 780
|||||
Db 721 gggaccgtgtgtctgtgacacactgaccttttttcgacctctctctgagacccacctg 780
QY 781 gaccagtccacggtgcatactcctcacacgcatactccacggcggtgtggtgtctccatg 840
Db 781 gaccagtccacggtgcatactcctcacacgcatactccacggcggtgtggtgtctccatg 840
QY 841 attctctggccttcacattattcttctgacctttctgacctttcccgaggaggttc 900
|||||
Db 841 attctctggccttcacattattcttctgacctttctgacctttcccgaggaggttc 900
QY 901 aagtcagaagatgcccaagaatccacgtggccctgggtggcagcctgttctcctgaat 960
Db 901 aagtcagaagatgcccaagaatccacgtggccctgggtggcagcctgttctcctgaat 960
QY 961 ctggccttcttgatgctggtggagtggtgctcaaaagggtctgagctgctgctgggcc 1020
Db 961 ctggccttcttgatgctggtggagtggtgctcaaaagggtctgagctgctgctgggcc 1020
QY 1021 cggggggctgtctcactactcctgtctgtctgtgcttcaactgagtgagccttgaagcc 1080
Db 1021 cggggggctgtctcactactcctgtctgtctgtgcttcaactgagtgagccttgaagcc 1080
QY 1081 ttccacctctacgtctgctgcaggggtcttcaacacctactctgggcactactctctg 1140
Db 1081 ttccacctctacgtctgctgcaggggtcttcaacacctactctgggcactactctctg 1140
QY 1141 aagctgagcctgtggtggctggccctgcccgcctgagtggtcagctggagtgcc 1200
Db 1141 aagctgagcctgtggtggctggccctgcccgcctgagtggtcagctggagtgcc 1200
QY 1201 aacagctacgctctacacctccgtgataggagaacgcacacctctctgagctatgc 1260
Db 1201 aacagctacgctctacacctccgtgataggagaacgcacacctctctgagctatgc 1260
QY 1261 tgggtccgtgaagggaacacatgtacgcccctctatatcacogtccacggctacttctc 1320
Db 1261 tgggtccgtgaagggaacacatgtacgcccctctatatcacogtccacggctacttctc 1320
QY 1321 ataccctctctttggcatggtggtcctggcctggtggtggaagatcttcaacctg 1380
Db 1321 ataccctctctttggcatggtggtcctggcctggtggtggaagatcttcaacctg 1380
QY 1381 tcccgctgctacagcgggtcaaggagcggggaagaaacccggaagaagtgctcaccctgctg 1440
Db 1381 tcccgctgctacagcgggtcaaggagcggggaagaaacccggaagaagtgctcaccctgctg 1440
QY 1441 ggcctctcgagcctgtggtggt 1461
Db 1441 ggcctctcgagcctgtgaaagt 1461

RESULT 4

AA080774
ID AA080774 standard; cDNA; 1458 BP.
XX AC AA080774;
XX DT 04-SEP-2001 (first entry)
XX

DE Human novel G-protein coupled receptor (NGPCR) #17.
XX Human; novel G-protein coupled receptor; NGPCR; signal transduction;
KW drug screening; transmembrane protein; mental disorder; gene therapy;
KW nootropic; pharmacogenomics; behavioural disorder; ss.
XX Homo sapiens.
FH Key Location/Qualifiers
FT CDS 1..1458
FT /*tag= a
FT /product= "Human NGPCR protein"
XX WO200142287-A2.
XX 14-JUN-2001.
XX 07-DEC-2000; 2000WO-US33241.
XX 07-DEC-1999; 99US-0169427.
XX (LEXI-) LEXICON GENETICS INC.
XX Donoho G, Scoville J, Turner CA, Friedrich G, Zambrowicz B;
PI Sands AT;
XX WPI: 2001-381634/40.
DR P-PSDB: AAE04464.
XX Novel polynucleotides encoding human G protein coupled receptors useful
for drug screening, diagnosis and in gene therapy of physiological or
behavioral disorders
XX Claim 5; Page 79; 91pp; English.
XX The present sequence is a cDNA encoding human novel G protein coupled
receptor (NGPCR). NGPCRs are transmembrane proteins that span the
cellular membrane and are involved in signal transduction after ligand
binding. NGPCR DNA sequences are useful as hybridisation probes for
screening libraries and assessing gene expression patterns. Sequences
derived from regions adjacent to the intron/exon boundaries of NGPCR
gene can be used to design primers for use in amplification assays to
detect mutations within the exons, splice sites, introns that can be
used in diagnostics and pharmacogenomics. NGPCR sequences and/or gene
products are useful for identifying compounds that modulate gene
expression or gene product activity. Such compounds are useful in the
treatment of mental, physiological or behavioural disorders and diseases.
CC NGPCR DNA sequences and antibodies are useful for diagnostic and
prognostic evaluation of disorders related to NGPCR function and for the
identification of subjects having a predisposition to such disorders.
CC NGPCR DNA sequences are also useful for drug screening and in gene
therapy for modulating NGPCR expression and to produce genetically
engineered host cells to express NGPCR products in vivo. The encoded
NGPCR proteins are useful for generating antibodies, as reagents in
diagnostic assays and for identifying other cellular gene products
related to NGPCR.
XX Sequence 1458 BP; 274 A; 442 C; 410 G; 332 T; 0 other;
SQ
Query Match 87.3%; Score 1441; DB 22; Length 1458;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 1458; Conservative 0; Mismatches 0; Indels 7; Gaps 1;
QY 1 atggcgagcggccaggggctggggccctgctctctgctctctctccacacaggt 60
|||||
Db 1 atggcgagcggccaggggctggggccctgctctctgctctctctccacacaggt 60
QY 61 caggaaaaagccacccaggggccaagaacacacctcctggggagcaacaacatgtacgac 120
|||||
Db 61 caggaaaaagccacccaggggccaagaacacacctcctggggagcaacaacatgtacgac 120
QY 121 atcttcaacttgatgacaaaggctttgtgttccaccagtgaggcaggtcggcagcgac 180

```

Db 121 |||||atcttcaacttgaaagacaagcttctgtcttccaccagtgaggcagctcgggacgac 180
Qy 181 |||||tctgcaatgttgaaaacttgacagatacttggtctaaactacgagggccatctgatgaag 240
Db 181 |||||tctgcaatgttgaaaacttgacagatacttggtctaaactacgagggccatctgatgaag 240
Qy 241 |||||gaaggtttgacgcagagaagtgaacacgctttctctgaagctttggtccagaacctaagc 300
Db 241 |||||gaaggtttgacgcagagaagtgaacacgctttctctgaagctttggtccagaacctaagc 300
Qy 301 |||||accaactgcagaagactctattctctctgagccctctcaggttccgagcaggtg 360
Db 301 |||||accaactgcagaagactctattctctctgagccctctcaggttccgagcaggtg 360
Qy 361 |||||atgaaggacgaggaacacccctgacagagtgcgacttcccaagagccttttccgaccc 420
Db 361 |||||atgaaggacgaggaacacccctgacagagtgcgacttcccaagagccttttccgaccc 420
Qy 421 |||||ctgcaggaacaagctgtgttcgcttgccgttgccgtccacattctggacattggtccaggg 480
Db 421 |||||ctgcaggaacaagctgtgttcgcttgccgttgccgtccacattctggacattggtccaggg 480
Qy 481 |||||actctctcaaggcccccggctggcctgggagatggcagcgctgtgtgaacaatcgc 540
Db 481 |||||actctctcaaggcccccggctggcctgggagatggcagcgctgtgtgaacaatcgc 540
Qy 541 |||||ctggtgggttgagtggtgggacaatgcattgcacaaagtggctgagcctctggagatc 600
Db 541 |||||ctggtgggttgagtggtgggacaatgcattgcacaaagtggctgagcctctggagatc 600
Qy 601 |||||gtctctctcaccagcagcccccataacatgacctcactgtgtattctgggagatg 660
Db 601 |||||gtctctctcaccagcagcccccataacatgacctcactgtgtattctgggagatg 660
Qy 661 |||||actaaagggaactggagactgtctcttgagggtgctccacggagggtcagacctgag 720
Db 661 |||||actaaagggaactggagactgtctcttgagggtgctccacggagggtcagacctgag 720
Qy 721 |||||gggacctgtgctgtgacacactgaccttttccgctgtctctgagaccacacttg 780
Db 721 |||||gggacctgtgctgtgacacactgaccttttccgctgtctctgagaccacacttg 780
Qy 781 |||||gaccagtccacggtgcatactccacacgactctccacggcggctgtgggtctccatg 840
Db 781 |||||gaccagtccacggtgcatactccacacgactctccacggcggctgtgggtctccatg 840
Qy 841 |||||atctctggccttcacacattattcttctgcttctgaggtttcccgaggaggttc 900
Db 841 |||||atctctggccttcacacattattcttctgcttctgaggtttcccgaggaggttc 900
Qy 901 |||||aagtcagaagatgccccaaagatcacagtgccctgggtggcagcctgttccctcctgaat 960
Db 901 |||||aagtcagaagatgccccaaagatcacagtgccctgggtggcagcctgttccctcctgaat 960
Qy 961 |||||ctgacctttgttgcatagtgggagtggtctcaaaagggtctgatgctgctgggcc 1020
Db 961 |||||ctgacctttgttgcatagtgggagtggtctcaaaagggtctgatgctgctgggcc 1020
Qy 1021 |||||cggggggctgtcttccactactctctctgtgcttctgacctgacctgaggtcagacc 1080
Db 1021 |||||cggggggctgtcttccactactctctgtgcttctgacctgacctgaggtcagacc 1080
Qy 1081 |||||ttcacactctactgctgctgctgaggttctcaacacactacttcggcactactctctg 1140
Db 1081 |||||ttcacactctactgctgctgctgaggttctcaacacactacttcggcactactctctg 1140
Qy 1141 |||||aagctgacctgtgggctggccctgcccgccctaatggttcacgactggagatgcc 1200
Db 1141 |||||aagctgacctgtgggctggccctgcccgccctaatggttcacgactggagatgcc 1200
Qy 1201 |||||aacagctacggcctctacacacatccctgtataggagaacccgacctctctggagctatgc 1260

```

```

Db 1201 |||||aacagctacggcctctacacacatccgtgtagggagaacgcacacctctctggagctatgc 1260
Qy 1261 |||||tggttcgctgaagggaacacacatgtagcgcctctctatacaccgtccacggctacttctc 1320
Db 1261 |||||tggttcgctgaagggaacacacatgtagcgcctctctatacaccgtccacggctacttctc 1320
Qy 1321 |||||ataccttctcttttggcatggtggtcctggcctggtggtggaagatcttcaacctg 1380
Db 1321 |||||ataccttctcttttggcatggtggtcctggcctggtggtggaagatcttcaacctg 1380
Qy 1381 |||||tccgctgctacacggtcgaagcagcgggggaagaaacgggaagagtgctcaccctgctg 1440
Db 1381 |||||tccgctgctacacggtcgaagcagcgggggaagaaacgggaagagtgctcaccctgctg 1440
Qy 1441 |||||ggcctctcagacctggtgggtgtga 1465
Db 1434 |||||ggcctctcagacctggtgggtgtga 1458

RESULT 5
AAD08764
ID AAD08764 standard; cDNA; 1515 bp.
XX
AC AAD08764;
XX
DT 04-SEP-2001 (first entry)
XX
DE Human novel G-protein coupled receptor (NGPCR) #7.
XX
KW Human; novel G-protein coupled receptor; NGPCR; signal transduction;
KW drug screening; transmembrane protein; mental disorder; gene therapy;
KW nootropic; pharmacogenomics; behavioural disorder; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..1515
FT /tag- a
FT /product= "Human NGPCR protein"
XX
PN WO200142287-A2.
XX
PD 14-JUN-2001.
XX
PF 07-DEC-2000; 2000WO-US33241.
XX
PR 07-DEC-1999; 99US-0169427.
XX
PA (LEXI-) LEXICON GENETICS INC.
XX
PI Donoho G, Scoville J, Turner CA, Friedrich G, Zambrowicz B;
PI Sands AT;
XX
DR WPI: 2001-381634/40.
DR P-PSDB: AAE04454.
XX
PT Novel polynucleotides encoding human G protein coupled receptors useful
PT for drug screening, diagnosis and in gene therapy of physiological or
PT behavioral disorders
XX
PS Disclosure: Page 67-68; 91pp; English.
XX
CC The present sequence is a cDNA encoding human novel G protein coupled
CC receptor (NGPCR). NGPCRs are transmembrane proteins that span the
CC cellular membrane and are involved in signal transduction after ligand
CC binding. NGPCR DNA sequences are useful as hybridisation probes for
CC screening libraries and assessing gene expression patterns. Sequences
CC derived from regions adjacent to the intron/exon boundaries of NGPCR
CC gene can be used to design primers for use in amplification assays to
CC detect mutations within the exons, splice sites, introns that can be
CC used in diagnostics and pharmacogenomics. NGPCR sequences and/or gene
CC products are useful for identifying compounds that modulate gene
CC expression or gene product activity. Such compounds are useful in the

```


XX Novel polynucleotides encoding human G protein coupled receptors useful
PT for drug screening, diagnosis and in gene therapy of physiological or
PT behavioral disorders -

XX Disclosure; Page 70; 91pp; English.

XX The present sequence is a cDNA encoding human novel G protein coupled
CC receptor (NGPCR). NGPCRs are transmembrane proteins that span the
CC cellular membrane and are involved in signal transduction after ligand
CC binding. NGPCR DNA sequences are useful as hybridisation probes for
CC screening libraries and assessing gene expression patterns. Sequences
CC derived from regions adjacent to the intron/exon boundaries of NGPCR
CC gene can be used to design primers for use in amplification assays to
CC detect mutations within the exons, splice sites, introns that can be
CC used in diagnostics and pharmacogenomics. NGPCR sequences and/or gene
CC products are useful for identifying compounds that modulate gene
CC expression or gene product activity. Such compounds are useful in the
CC treatment of mental, physiological or behavioural disorders and diseases.
CC NGPCR DNA sequences and antibodies are useful for diagnostic and
CC prognostic evaluation of disorders related to NGPCR function and for the
CC identification of subjects having a predisposition to such disorders.
CC NGPCR DNA sequences are also useful for drug screening and in gene
CC therapy for modulating NGPCR expression and to produce genetically
CC engineered host cells to express NGPCR products in vivo. The encoded
CC NGPCR proteins are useful for generating antibodies, as reagents in
CC diagnostic assays and for identifying other cellular gene products
CC related to NGPCR.

XX SQ Sequence 1068 BP; 177 A; 336 C; 298 G; 257 T; 0 other;

Query Match 57.1%; Score 941.6; DB 22; Length 1068;

Best Local Similarity 98.4%; Pred. No. 1.3e-221;
Matches 964; Conservative 0; Mismatches 9; Indels 7; Gaps 1;

QY 482 ctctctcaaggcccccgcgctcggtgagatggcagcggtgttgaaacaatgcc 541
DB 35 ccccttcacaggcccccgcgctcggtgagatggcagcggtgttgaaacaatgcc 94
QY 542 tgggtgggttgagtgaggacaaaatgcatgtcaccaagctggctgagcctctggagatcg 601
DB 95 tgggtgggttgagtgaggacaaaatgcatgtcaccaagctggctgagcctctggagatcg 154
QY 602 tcttctctaccagcagcccccctaaatgacacctcaactgtgtattctggagatga 661
DB 155 tcttctctaccagcagcccccctaaatgacacctcaactgtgtattctggagatga 214
QY 662 ctaaaggaccactggagactggtcttctgagggtgtctccacgaggtcagacactgag 721
DB 215 ctaaaggaccactggagactggtcttctgagggtgtctccacgaggtcagacactgag 274
QY 722 ggaccgtgtgtgtgtgaccacactgacacttttgcgctgtctctgagaccacactgg 781
DB 275 ggaccgtgtgtgtgtgaccacactgacacttttgcgctgtctctgagaccacactgg 334
QY 782 accagtcacaggtgcatctcctcaacgcatctccacgagcggtgtgggtctccatga 841
DB 335 accagtcacaggtgcatctcctcaacgcatctccacgagcggtgtgggtctccatga 394
QY 842 tcttctggccttcacattattcttctgaggtgttcccgaggaggttca 901
DB 395 tcttctggccttcacattattcttctgaggtgttcccgaggaggttca 454
QY 902 agtcagaagatgcccaagaatccacgtggccctgggtggcagcctgttccctcgaatc 961
DB 455 agtcagaagatgcccaagaatccacgtggccctgggtggcagcctgttccctcgaatc 514
QY 962 tggccttcttggccaatgtggggagtggtcacaaggggtctgatgctgctgtgggcc 1021
DB 515 tggccttcttggccaatgtggggagtggtcacaaggggtctgatgctgctgtgggcc 574
QY 1022 ggggggctgtcttccacttctctgtctgtgcttccactggatggggccttgagcct 1081

Db 575 ggggggctgtcttccacttctctgtctgtgcttccactggatgggcttgaagcct 634
QY 1082 tccacctctactctcgtcgtgcaagggtcttcaacactacttccggcactacttctga 1141
Db 635 tccacctctactctcgtcgtgcaagggtcttcaacactacttccggcactacttctga 694
QY 1142 agctgagcctggtgggctgcccgcctgacccgcctgatggtcatcgccactgggagtgcca 1201
Db 695 agctgagcctggtgggctgcccgcctgacccgcctgatggtcatcgccactgggagtgcca 754
QY 1202 acagtctacggcctctacacatccgtgataggaggagaaccacactctctggagctatgct 1261
Db 755 acagtctacggcctctacacatccgtgataggaggagaaccacactctctggagctatgct 814
QY 1262 gtttcgtgaaggagaacacatgtacgcctctctatacacgctcacgctacttctca 1321
Db 815 gtttcgtgaaggagaacacatgtacgcctctctatacacgctcacgctacttctca 874
QY 1322 tcaccttctcttggcatggtgctctgcccctggtggtctggaagatcttccacctgt 1381
Db 875 tcaccttctcttggcatggtgctctgcccctggtggtctggaagatcttccacctgt 934
QY 1382 cccgtgtctacagcgtcaagagcgggggaagaaacccggaagaggtgtctcaccctgtg 1441
Db 935 cccgtgtctacagcgtcaagagcgggggaagaaacccggaagaggtgtctcaccctgtg 987
QY 1442 gcctctgagcctgtgtgggt 1461
Db 988 gcctctgagcctgtgcaagt 1007

RESULT 12

AAD08770
ID AAD08770 standard; cDNA; 936 BP.

XX AAD08770;
XX 04-SEP-2001 (first entry)
XX Human novel G-protein coupled receptor (NGPCR) #13.
XX Human; novel G-protein coupled receptor; NGPCR; signal transduction;
KW drug screening; transmembrane protein; mental disorder; gene therapy;
KW nootropic; pharmacogenomics; behavioural disorder; ss.
XX Homo sapiens.
XX Key Location/Qualifiers
FT CDS 1..936
FT /tag- a
FT /product= "Human NGPCR protein"
XX WO200142287-A2.
XX 14-JUN-2001.
XX 07-DEC-2000; 2000WO-US33241.
XX 07-DEC-1999; 99US-0169427.
XX (LEXI-) LEXICON GENETICS INC.
XX Donoho G, Scoville J, Turner CA, Friedrich G, Zambrowicz B;
PI Sands AT;
XX WPI; 2001-381634/40.
XX P-PSDB; AAE04460.
XX Novel polynucleotides encoding human G protein coupled receptors useful
PT for drug screening, diagnosis and in gene therapy of physiological or
PT behavioral disorders -
XX

PS Disclosure; Page 75; 91pp; English.

CC The present sequence is a cDNA encoding human novel G protein coupled

CC receptor (NGPCR). NGPCRs are transmembrane proteins that span the

CC cellular membrane and are involved in signal transduction after ligand

CC binding. NGPCR DNA sequences are useful as hybridisation probes for

CC screening libraries and assessing gene expression patterns. Sequences

CC derived from regions adjacent to the intron/exon boundaries of NGPCR

CC gene can be used to design primers for use in amplification assays to

CC detect mutations within the exons, splice sites, introns that can be

CC used in diagnostics and pharmacogenomics. NGPCR sequences and/or gene

CC expression or gene product activity. Such compounds are useful in the

CC treatment of mental, physiological or behavioural disorders and diseases.

CC NGPCR DNA sequences and antibodies are useful for diagnostic and

CC prognostic evaluation of disorders related to NGPCR function and for the

CC identification of subjects having a predisposition to such disorders.

CC NGPCR DNA sequences are also useful for drug screening and in gene

CC therapy for modulating NGPCR expression and to produce genetically

CC engineered host cells to express NGPCR products in vivo. The encoded

CC NGPCR proteins are useful for generating antibodies, as reagents in

CC diagnostic assays and for identifying other cellular gene products

CC related to NGPCR.

XX

SQ

Sequence 936 BP; 153 A; 289 C; 263 G; 231 T; 0 other;

Query Match 55.3%; Score 913; DB 22; Length 936;
Best Local Similarity 100.0%; Pred. No. 1.3e-214;
Matches 913; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 631 atgacctcactgtattctggtggtgactaaaggagaccactgagactggtctct 690

Db 1 atgacctcactgtattctggtggtgactaaaggagaccactgagactggtctct 60

Qy 691 gagggctgtccacgaggttcagacctgagggaacctgtgctgtgaccacctgacc 750

Db 61 gagggctgtccacgaggttcagacctgagggaacctgtgctgtgaccacctgacc 120

Qy 751 ttttgcctgtccttgagaccaccttgaccagtcacagtcacagtcacacacgc 810

Db 121 ttttgcctgtccttgagaccaccttgaccagtcacagtcacagtcacacacgc 180

Qy 811 atctccacgagggctgtggtctcctgactgactgcttccctcacttattcttat 870

Db 181 atctccacgagggctgtggtctcctgactgactgcttccctcacttattcttat 240

Qy 871 gctttctgaggtttcccgaggaggttcaagtcaagaagatgcccaagatccacgtg 930

Db 241 gctttctgaggtttcccgaggaggttcaagtcaagaagatgcccaagatccacgtg 300

Qy 931 gctctgggtggcagcctgttctcctgtaacttgaccttcttgctcaatgtggggagtgc 990

Db 301 gctctgggtggcagcctgttctcctgtaacttgaccttcttgctcaatgtggggagtgc 360

Qy 991 tcaaaagggtctatgctcctgactgagccggccgggggctgtcttccactacttctgctc 1050

Db 361 tcaaaagggtctatgctcctgactgagccggccgggggctgtcttccactacttctgctc 420

Qy 1051 tgtgcttcaacctggatggccttgagccttccaccttccaccttaccgtcgtcgaggttc 1110

Db 421 tgtgcttcaacctggatggccttgagccttccaccttccaccttaccgtcgtcgaggttc 480

Qy 1111 ttaaacaccttactcggcacttctcctgaagctgagcctggtgggctgggctgcgc 1170

Db 481 ttaaacaccttactcggcacttctcctgaagctgagcctggtgggctgggctgcgc 540

Qy 1171 gctctgatgtcactcagcactgaggtgccaacagactcagcctctacacactcctgat 1230

Db 541 gctctgatgtcactcagcactgaggtgccaacagactcagcctctacacactcctgat 600

Qy 1231 agggagaacgcacctctctcctgagctatgctggttccgtgagggagacacactgtacgc 1290

Db 601 agggagaacgcacctctctcctgagctatgctggttccgtgaaggagacacactgtacgc 660

Qy 1291 ctctatatcaccgtccacagcctacttctcctcactcactccttcttctgcatgggtgctcg 1350

Db 661 ctctatatcaccgtccacagcctacttctcctcactcactccttcttctgcatgggtgctcg 720

Qy 1351 gctctggtgctggaagatcttccacctgctccctgctcagcgtgctcagagcgtgagcgtg 1410

Db 721 gctctggtgctggaagatcttccacctgctccctgctcagcgtgctcagagcgtgagcgtg 780

Qy 1411 aagaacccgagaaggtgctcaccctgctggcctcctcagcgtgctggtgtgacatgg 1470

Db 781 aagaacccgagaaggtgctcaccctgctggcctcctcagcgtgctggtgtgacatgg 840

Qy 1471 ggggttgccattctcaccctggtggcctcctcaccgtctacatcttgcacttttcaac 1530

Db 841 ggggttgccattctcaccctggtggcctcctcaccgtctacatcttgcacttttcaac 900

Qy 1531 tccctgcaagggtg 1543

Db 901 tccctgcaagggtg 913

RESULT 13

AAD08760

ID AAD08760 standard; cDNA; 897 BP.

XX

AC AAD08760;

XX

DT 04-SEP-2001 (first entry)

XX

DE Human novel G-protein coupled receptor (NGPCR) #3.

XX

KW Human; novel G-protein coupled receptor; NGPCR; signal transduction;

KW drug screening; transmembrane protein; mental disorder; gene therapy;

KW neurotic; pharmacogenomics; behavioural disorder; ss.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT CDS 1..897

FT /tag= a

FT /product= "Human NGPCR protein"

XX

PN WO200142287-A2.

XX

PD 14-JUN-2001.

XX

PF 07-DEC-2000; 2000WO-US33241.

XX

PR 07-DEC-1999; 99US-0169427.

XX

PA (LEXI-) LEXICON GENETICS INC.

XX

PI Donoho G, Scoville J, Turner CA, Friedrich G, Zambrowicz B;

PI Sands AT;

XX

DR WPI; 2001-381634/40.

XX

DR P-PSDB; AAE04450.

XX

PT Novel polynucleotides encoding human G protein coupled receptors useful

PT for drug screening, diagnosis and in gene therapy of physiological or

PT behavioral disorders -

XX

PS Disclosure; Page 64; 91pp; English.

XX

CC The present sequence is a cDNA encoding human novel G protein coupled

CC receptor (NGPCR). NGPCRs are transmembrane proteins that span the

CC cellular membrane and are involved in signal transduction after ligand

CC binding. NGPCR DNA sequences are useful as hybridisation probes for

CC screening libraries and assessing gene expression patterns. Sequences

CC derived from regions adjacent to the intron/exon boundaries of NGPCR

CC gene can be used to design primers for use in amplification assays to

CC

Query Match 49.2%; Score 811; DB 22; Length 828;
Best Local Similarity 99.2%; Pred. No. 1.5e-189;
Matches 828; Conservative 0; Mismatches 0; Indels 7; Gaps 1;

QY 631 atgacctcaactgtgtattctggatgtgactaaaggagaccactggagactggtcttct 690
Db 1 atgacctcaactgtgtattctggatgtgactaaaggagaccactggagactggtcttct 60

QY 691 gagggtctctcaacaggaggtcagacctgaggggacggtgtgctgtgacacactgacc 750
Db 61 gagggtctctcaacaggaggtcagacctgaggggacggtgtgctgtgacacactgacc 120

QY 751 ttttcgcccgtctctcagaccaccttgaccaggtccacgggtgcatactcaccacgc 810
Db 121 ttttcgcccgtctctcagaccaccttgaccaggtccacgggtgcatactcaccacgc 180

QY 811 atctccaggcggcgtggtggtctccatgatctctctcgtcctcaccattattcttat 870
Db 181 atctccaggcggcgtggtggtctccatgatctctcgtcctcaccattattcttat 240

QY 871 gcccttctgagcgttcccgaggaggtcgaagtcaagatgcccaagatccacgtg 930
Db 241 gcccttctgagcgttcccgaggaggtcgaagtcaagatgcccaagatccacgtg 300

QY 931 gccctggtgagcgttccctcctgaatcgtccttcttctggtcgaatgtggaggtgc 990
Db 301 gccctggtgagcgttccctcctgaatcgtccttcttctggtcgaatgtggaggtgc 360

QY 991 tcaagggtgtgctgctgctggtggcggggcggcgtgcttccactacttctctc 1050
Db 361 tcaagggtgtgctgctgctggtggcggggcggcgtgcttccactacttctctc 420

QY 1051 tgtgcttcaactgagtgaggccttgaggccttccacctctacctgtcgtctcagggtc 1110
Db 421 tgtgcttcaactgagtgaggccttgaggccttccacctctacctgtcgtctcagggtc 480

QY 1111 tcaacactactcggcactactcctgaagctgagcctggtggtggcctggcctgcc 1170
Db 481 tcaacactactcggcactactcctgaagctgagcctggtggtggcctggcctgcc 540

QY 1171 gccctgagtgatcggcactggagtgcccaacagctacgctctacacatccgtgat 1230
Db 541 gccctgagtgatcggcactggagtgcccaacagctacgctctacacatccgtgat 600

QY 1231 agggagaacgcacctctgagcactggtggttcggtgaaggagacacacatgacgcc 1290
Db 601 agggagaacgcacctctgagcactggtggttcggtgaaggagacacacatgacgcc 660

QY 1291 ctctataccgtccacggtcactctcctcatcacttctcttcttggcatggtgctctg 1350
Db 661 ctctataccgtccacggtcactctcctcatcacttctcttcttggcatggtgctctg 720

QY 1351 gccctggtgtctggaagatcttcaacctgtcccgctgctacagcgtcgaaggagcgggg 1410
Db 721 gccctggtgtctggaagatcttcaacctgtcccgctgctacagcgtcgaaggagcgggg 780

QY 1411 aagaacgcgaagaggtgtcaccctgtcggcctctcagcctgtggtgtga 1465
Db 781 aagaac-----ggtgtcaccctgtcggcctctcagcctgtggtgtga 828

RESULT 15
AAD08765
ID AAD08765 standard; cDNA; 885 BP.
XX AAD08765;
AC AAD08765;
DT 04-SEP-2001 (first entry)
XX Human novel G-protein coupled receptor (NGPCR) #8.
DE Human; novel G-protein coupled receptor; NGPCR; signal transduction;
KW Human; novel G-protein coupled receptor; NGPCR; signal transduction;

KW drug screening; transmembrane protein; mental disorder; gene therapy;
KW nootropic; pharmacogenomics; behavioural disorder; ss.
XX Homo sapiens.
XX Location/Qualifiers
FH 1..885
FT /*tag= a
FT /product= "Human NGPCR protein"
XX WO200142287-A2.
XX 14-JUN-2001.
XX 07-DEC-2000; 2000WO-US33241.
XX 07-DEC-1999; 99US-0169427.
XX (LEXI-) LEXICON GENETICS INC.
XX Donoho G, Scoville J, Turner CA, Friedrich G, Zambrowicz B;
XX Sands AT;
XX WPI; 2001-381634/40.
XX P-PSDB; AAE04455.
XX Novel polynucleotides encoding human G protein coupled receptors useful
XX for drug screening, diagnosis and in gene therapy of physiological or
XX behavioral disorders.
XX Disclosure; Page 69; 91pp; English.
XX The present sequence is a cDNA encoding human novel G protein coupled
XX receptor (NGPCR). NGPCRs are transmembrane proteins that span the
XX cellular membrane and are involved in signal transduction after ligand
XX binding. NGPCR DNA sequences are useful as hybridisation probes for
XX screening libraries and assessing gene expression patterns. Sequences
XX derived from regions adjacent to the intron/exon boundaries of NGPCR
XX gene can be used to design primers for use in amplification assays to
XX detect mutations within the exons, splice sites, introns that can be
XX used in diagnostics and pharmacogenomics. NGPCR sequences and/or gene
XX products are useful for identifying compounds that modulate gene
XX expression or gene product activity. Such compounds are useful in the
XX treatment of mental, physiological or behavioural disorders and diseases.
XX NGPCR DNA sequences and antibodies are useful for diagnostic and
XX prognostic evaluation of disorders related to NGPCR function and for the
XX identification of subjects having a predisposition to such disorders.
XX NGPCR DNA sequences are also useful for drug screening and in gene
XX therapy for modulating NGPCR expression and to produce genetically
XX engineered host cells to express NGPCR products in vivo. The encoded
XX NGPCR proteins are useful for generating antibodies, as reagents in
XX diagnostic assays and for identifying other cellular gene products
XX related to NGPCR.
XX Sequence 885 BP; 148 A; 273 C; 246 G; 218 T; 0 other;

Query Match 48.5%; Score 800.6; DB 22; Length 885;
Best Local Similarity 98.7%; Pred. No. 5.6e-187;
Matches 820; Conservative 0; Mismatches 4; Indels 7; Gaps 1;

QY 631 atgacctcaactgtgtattctggatgtgactaaaggagaccactggagactggtcttct 690
Db 1 atgacctcaactgtgtattctggatgtgactaaaggagaccactggagactggtcttct 60

QY 691 gagggtctctcaacaggaggtcagacctgaggggacggtgtgctgtgacacactgacc 750
Db 61 gagggtctctcaacaggaggtcagacctgaggggacggtgtgctgtgacacactgacc 120

QY 751 ttttcgcccgtctctcagaccaccttgaccaggtccacgggtgcatactcaccacgc 810
Db 121 ttttcgcccgtctctcagaccaccttgaccaggtccacgggtgcatactcaccacgc 180

```

QY 811 atctccaggcggcgtgtggggtccatgatcttctcgtcccttccaccattattttat 870
Db |||||||
QY 181 atctccaggcggcgtgtggggtcccatgatcttctcgtcccttccaccattattttat 240
Db |||||||
QY 871 gcctttctgagcgtttcccgaggaggttcaagtccagaagatgccccaaagatccacgtg 930
Db |||||||
QY 241 gcctttctgagcgtttcccgaggaggttcaagtccagaagatgccccaaagatccacgtg 300
Db |||||||
QY 931 gccttgggtggcagcgtgttctctcgtgaatctggccttcttgggtcaatgtgggagtgcc 990
Db |||||||
QY 301 gccttgggtggcagcgtgttctctcgtgaatctggccttcttgggtcaatgtgggagtgcc 360
Db |||||||
QY 991 tcaaaagggtctgagctgcctgcctggcccggggggctgtcttccactacttctcctc 1050
Db |||||||
QY 361 tcaaaagggtctgagctgcctgcctggcccggggggctgtcttccactacttctcctc 420
Db |||||||
QY 1051 tgtgccttcacctggatggccttgaagccttccacccttccactcctcgtcgtcagggtc 1110
Db |||||||
QY 421 tgtgccttcacctggatggccttgaagccttccacccttccactcctcgtcgtcagggtc 480
Db |||||||
QY 1111 ttaaacactacttccggcactacttctgaagctgagcctggtggcgtggcctgcc 1170
Db |||||||
QY 481 ttaaacactacttccggcactacttctgaagctgagcctggtggcgtggcctgcc 540
Db |||||||
QY 1171 gcctgatgtcatcggcactgggagtgccaacagctacggcctctacaccatccgtgat 1230
Db |||||||
QY 541 gcctgatgtcatcggcactgggagtgccaacagctacggcctctacaccatccgtgat 600
Db |||||||
QY 1231 agggagaaccgcacctctctgagctatgtcgttccgtgaaggagacaaccatgtacgcc 1290
Db |||||||
QY 601 agggagaaccgcacctctctgagctatgtcgttccgtgaaggagacaaccatgtacgcc 660
Db |||||||
QY 1291 ctctatatcccgccacggctacttctcctacatcccttcttggcatggtgctcctg 1350
Db |||||||
QY 661 ctctatatcccgccacggctacttctcctacatcccttcttggcatggtgctcctg 720
Db |||||||
QY 1351 gccttgggtgtctggaagatcttccacctgtcccggtgctacacggtcaaggagcggggg 1410
Db |||||||
QY 721 gccttgggtgtctggaagatcttccacctgtcccggtgctacacggtcaaggagcggggg 780
Db |||||||
QY 1411 aagaaccggaagaaggtgtcaccctgtcggcctctcagcctgtggtgggt 1461
Db |||||||
QY 781 aagaacc-----ggtgctcaccctgtcggcctctcagcctgtgcaagt 824

```

Search completed: June 11, 2002, 21:31:16
Job time: 4275 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 11, 2002, 20:17:59 ; Search time 53.57 Seconds
(without alignments)
7565.714 Million cell updates/sec

Title: US-09-733-387-43
Perfect score: 1650
Sequence: 1 atggagagccagggcct.....actcgcattcagaataag 1650

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues
Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description
1	145.8	8.8	3090	4	US-09-276-531-78
2	85.2	5.2	669	4	US-09-276-531-125
3	51.6	3.1	7218	1	US-08-232-463-14
4	48.6	2.9	1931	2	US-09-130-114-2
5	46.2	2.8	1027	4	US-09-276-531-24
6	45.4	2.8	12588	2	US-08-387-942C-1
7	43.6	2.6	1521	1	US-08-083-948-14
8	43.6	2.6	1521	1	US-08-393-785-14
9	43.6	2.6	1521	1	US-08-475-694-14
10	43.6	2.6	1521	1	US-08-712-057-14
11	43.6	2.6	1647	1	US-08-083-948-13
12	43.6	2.6	1647	1	US-08-393-785-13
13	43.6	2.6	1647	1	US-08-475-694-13
14	43.6	2.6	1647	1	US-08-712-057-13
15	43.6	2.6	1865	1	US-08-083-948-7
16	43.6	2.6	1865	1	US-08-393-785-7
17	43.6	2.6	1865	1	US-08-475-694-7
18	43.6	2.6	1865	1	US-08-712-057-7
19	43.4	2.6	1468	1	US-08-381-433A-7
20	43.2	2.6	459	2	US-08-387-942C-31
21	43.2	2.6	5894	3	US-08-665-259-24
22	43.2	2.6	5894	3	US-08-762-500-24
23	43.2	2.6	6525	3	US-08-762-500-74
24	42	2.5	1288	1	US-08-440-856A-9
25	41.6	2.5	3058	4	US-09-276-531-107
26	40.8	2.5	1893	3	US-09-320-878-23
27	40.6	2.5	975	4	US-09-365-150-3
28					Sequence 78, Appl
29					Sequence 125, Appl
30					Sequence 14, Appl
31					Sequence 2, Appl
32					Sequence 24, Appl
33					Sequence 1, Appl
34					Sequence 14, Appl
35					Sequence 14, Appl
36					Sequence 14, Appl
37					Sequence 13, Appl
38					Sequence 13, Appl
39					Sequence 7, Appl
40					Sequence 7, Appl
41					Sequence 7, Appl
42					Sequence 7, Appl
43					Sequence 31, Appl
44					Sequence 24, Appl
45					Sequence 74, Appl
					Sequence 9, Appl
					Sequence 107, Appl
					Sequence 23, Appl
					Sequence 3, Appl

28	40.2	2.4	1862	1	US-07-864-475A-1	Sequence 1, Appl
29	40.2	2.4	1862	2	US-08-468-249A-1	Sequence 2, Appl
30	40.2	2.4	1863	1	US-07-864-475A-2	Sequence 2, Appl
31	40.2	2.4	1863	2	US-08-468-249A-2	Sequence 1, Appl
32	40.2	2.4	3271	2	US-08-852-806-1	Sequence 1, Appl
33	40.2	2.4	3271	3	US-09-163-669-1	Sequence 1, Appl
34	39.8	2.4	756	3	US-08-513-974B-53	Sequence 53, Appl
35	39.8	2.4	810	3	US-08-086-439C-2	Sequence 359, Appl
36	39.8	2.4	1161	1	US-08-434-877-2	Sequence 2, Appl
37	39.8	2.4	1161	1	US-08-434-877-2	Sequence 2, Appl
38	39.8	2.4	1367	3	US-08-475-742-3	Sequence 3, Appl
39	39.6	2.4	390	4	US-09-197-649-7	Sequence 7, Appl
40	39.2	2.4	5698	1	US-08-761-258-11	Sequence 11, Appl
41	39.2	2.4	5698	2	US-08-977-306-11	Sequence 11, Appl
42	38.6	2.3	2150	2	US-08-861-464-13	Sequence 13, Appl
43	38.6	2.3	2150	3	US-08-396-001-13	Sequence 13, Appl
44	38.6	2.3	2150	4	US-09-323-433A-13	Sequence 13, Appl
45	38.6	2.3	2207	4	US-08-956-322-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1
US-09-276-531-78
; Sequence 78, Application US/09276531
; Patent No. 6183968
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Yue, Henry
; APPLICANT: Reddy, Roopa
; APPLICANT: Guegler, Karl J.
; APPLICANT: Baughn, Mariah R.
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF GENES ENCODING
; TITLE OF INVENTION: RECEPTORS AND PROTEINS ASSOCIATED WITH CELL PROLIFERATION
; NUMBER OF SEQUENCES: 134
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/276,531
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/079,677
; FILING DATE: March 27, 1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Lynn E. Murry, Ph.D.
; REGISTRATION NUMBER: 42,918
; REFERENCE/DOCKET NUMBER: PA-0008 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 853-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 78:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3090 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BRAINOT14

```

: APPLICANT: Baughn, Mariah R.
:
: TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF GENES ENCODING
:
: TITLE OF INVENTION: RECEPTORS AND PROTEINS ASSOCIATED WITH CELL PROLIFERATION
:
: NUMBER OF SEQUENCES: 134
:
: CORRESPONDENCE ADDRESS:
:

```

; NUMBER OF SEQUENCES: 52

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZgpt-F1s
US-08-232-463-14

Query Match 3.1%; Score 51.6; DB 1; Length 7218;
Best Local Similarity 11.7%; Pred. No. 0.0016;
Matches 57; Conservative 211; Mismatches 220; Indels 0; Gaps 0;

QY 902 agtcagaagatgcccaagatcacgctggtccctgggtggcagcctgttccctcgaatc 961
Db 934 ATTCAAATAATATTAAACGGTTAGTTGAATGTCCTATCGAGTCCGCTACTATAAC 993
QY 962 tggcctcttgggtcaatgtgggagtggtcgaagggtgtgatgctgctgctggccc 1021
Db 994 TATTTTTCCTCGTTTGCCATACGCTCACAGAATTAATTCGAGCTTGCTGCAGTCCA 1053
QY 1022 ggggggctgtctccactactctctgctgtgctctcactcctgagtggtcctgaagct 1081
Db 1054 GGGAGCTTGCATYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1113
QY 1082 tcacctctactctgctgctgaggtcttcaacacttctcgggcactactctctga 1141
Db 1114 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1173
QY 1142 agctgagcctgggtgggctgcccgcctcctgctgctgctgctgctgctgctgctg 1201
Db 1174 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1233
QY 1202 acagctacgctctacacatccgtgatagggagaaacgcacactctctgagactatgct 1261
Db 1234 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1293
QY 1262 ggtccctggaaggacaacatgacgcctctatcacctcctcctcctcctcctcctc 1321
Db 1294 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1353
QY 1322 tcacctctcttggcagtggtgctgctgctgctgctgctgctgctgctgctgctgct 1381

CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZgpt-F1s
US-08-232-463-14

Query Match 2.9%; Score 48.6; DB 2; Length 1931;
Best Local Similarity 43.0%; Pred. No. 0.0052;
Matches 237; Conservative 0; Mismatches 314; Indels 0; Gaps 0;

QY 1030 gtcttcactactctctgctgctgctcactcctggtggtggtggtggtggtggtg 1089
Db 299 gtctcccccgtctctgctctcctcctcctcctcctcctcctcctcctcctcctc 358
QY 1090 tacctgctgctgctcaggtgttcaaacacttactcgggacactctctcctgaaactgagc 1149
Db 359 cccgtctctcccgctgctgctcctcctcctcctcctcctcctcctcctcctcctc 418
QY 1150 ttgtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 1209
Db 419 ctccccctgctgctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 478
QY 1210 ggcctctacacatccgtgataggaggaacgcacactctctggagctatgctggttcgt 1269
Db 479 ctctctcccgctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 538
QY 1270 gaagggacaacatgtagccctctatatcccgctccacggctacttctctcactcactc 1329
Db 539 cccgtctcccgctgctgctcctcctcctcctcctcctcctcctcctcctcctcctc 598
QY 1330 ctctttgcatggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 1389
Db 599 ctccccctgctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 658
QY 1390 acagcgttcaagagcggggggaagaaacgcagagaggtgctcaccctgctggtcctcg 1449
Db 659 gtctctcccgctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 718
QY 1450 agcctggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 1509
Db 719 gtctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 778
QY 1510 tacatcttgcacttttaactccttgcaagggtgcttcttcttcttcttcttcttct 1569
Db 779 cccgtcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 838
QY 1570 ctttaacctccc 1580
Db 839 gtctcctccc 849
```


[illegible]

```

RESULT      8
US-08-393-785-14
; Sequence 14, Application US/08393785
; Patent No. 5554369
; GENERAL INFORMATION:
; APPLICANT: Corbin, David R.
; APPLICANT: Greenplate, John T.
; APPLICANT: Jennings, Michael G.
; APPLICANT: Purcell, John P.
; APPLICANT: Sammons, Robert D.
; TITLE OF INVENTION: Method of Controlling Insects
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Grace L. Bonner, Monsanto Company, B84F
; STREET: 700 Chesterfield Parkway No. 5554369th
; CITY: St. Louis
; STATE: Missouri
; COUNTRY: USA
; ZIP: 63198
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/393,785
; FILING DATE: 24-FEB-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/083,948
; FILING DATE: 28-JUN-1993
; APPLICATION NUMBER: US 07/762682
; FILING DATE: 23-SEP-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/937195
; FILING DATE: 09-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Bonner, Grace L.

```


CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/762682
FILING DATE: 23-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937195
FILING DATE: 09-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Bonner, Grace L.
REGISTRATION NUMBER: 32,963
REFERENCE/DOCKET NUMBER: 38-21(10631)A
TELEPHONE: (314)537-7286
TELEFAX: (314)537-6047
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 1521 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-712-057-14

Query Match 2.6%; Score 43.6; DB 1; Length 1521;
Best Local Similarity 47.2%; Pred. No. 0.086;
Matches 133; Conservative 0; Mismatches 149; Indels 0; Gaps 0;

QY 1109 tcttaaacacacttctggcacttctgaagctgagcctggtggcctgc 1168
Db 845 TCTTCTCGGCGCGGCGAGCTCGCTCCATTGAACGTGCTGGCGCGCGGAGACGG 904
QY 1169 ccgacctgattgctcggcactgggagtgccaaacagctacgctctacacaccctgc 1228
Db 905 GCACCTGCGCGCGCTCAGCTCGAGATCGGCGCGCTGGGGCCCCAACGCAACATCA 964
QY 1229 ataggagaaacgcacactctctgagctatgtgttcctggaaggacacacatgtacg 1288
Db 965 TGACCGCGCGCGCGCAACCATGTGTGGAACCCGACGCGGAGCAAGCAGTGTGATCCCG 1024
QY 1289 cctctatatcacgctccacgctacttctctcaccctgtcccgctgcta 1348
Db 1025 CCCTCGGATCGAGACTGGGAGAACCCGACAAACCCGCTTCGCGGAGATAGCCCCA 1084
QY 1349 tggccctgtggtctggaagattctcaccctgtcccgctgcta 1390
Db 1085 TGCGCGCGCGCTCGAGACCTGGGTGAGCTTACCTGSCCA 1126

RESULT 11
US-08-083-948-13
Sequence 13, Application US/08083948
Patent No. 5518908
GENERAL INFORMATION:
APPLICANT: Corbin, David R.
APPLICANT: Greenplate, John T.
APPLICANT: Jennings, Michael G.
APPLICANT: Purcell, John P.
APPLICANT: Sammons, Robert D.
TITLE OF INVENTION: Method of Controlling Insects
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Grace L. Bonner, Monsanto Company, BB4F
STREET: 700 Chesterfield Parkway No. 5518908th
CITY: St. Louis
STATE: Missouri
COUNTRY: USA
ZIP: 63198
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/083,948
FILING DATE: 19930628
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/762682
FILING DATE: 23-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937195
FILING DATE: 09-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Bonner, Grace L.
REGISTRATION NUMBER: 32,963
REFERENCE/DOCKET NUMBER: 38-21(10631)A
TELEPHONE: (314)537-7286
TELEFAX: (314)537-6047
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 1647 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-083-948-13

Query Match 2.6%; Score 43.6; DB 1; Length 1647;
Best Local Similarity 47.2%; Pred. No. 0.089;
Matches 133; Conservative 0; Mismatches 149; Indels 0; Gaps 0;

QY 1109 tcttaaacacacttctggcacttctgaagctgagcctggtggcctgc 1168
Db 971 TCTTCTCGGCGCGGCGAGCTCGCTCCATTGAACGTGCTGGCGCGCGGAGACGG 1030
QY 1169 ccgacctgattgctcggcactgggagtgccaaacagctacgctctacacaccctgc 1228
Db 1031 GCACCTGCGCGCGCTCAGCTCGAGATCGGCGCGCTGGGGCCCCAACGCAACATCA 1090
QY 1229 ataggagaaacgcacactctctgagctatgtgttcctggaaggacacacatgtacg 1288
Db 1091 TGACCGCGCGCGCGCAACCATGTGTGGAACCCGACGCGGAGCAAGCAGTGTGATCCCG 1150
QY 1289 cctctatatcacgctccacgctacttctctcaccctgtcccgctgcta 1348
Db 1151 CCCTCGGATCGAGACTGGGAGAACCCGACAAACCCGCTTCGCGGAGATAGCCCCA 1210
QY 1349 tggccctgtggtctggaagattctcaccctgtcccgctgcta 1390
Db 1211 TGCGCGCGCGCTCGAGACCTGGGTGAGCTTACCTGSCCA 1252

RESULT 12
US-08-393-785-13
Sequence 13, Application US/08393785
Patent No. 5554369
GENERAL INFORMATION:
APPLICANT: Corbin, David R.
APPLICANT: Greenplate, John T.
APPLICANT: Jennings, Michael G.
APPLICANT: Purcell, John P.
APPLICANT: Sammons, Robert D.
TITLE OF INVENTION: Method of Controlling Insects
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Grace L. Bonner, Monsanto Company, BB4F
STREET: 700 Chesterfield Parkway No. 5554369th
CITY: St. Louis
STATE: Missouri
COUNTRY: USA
ZIP: 63198
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/393,785
; FILING DATE: 24-FEB-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/083,948
; FILING DATE: 28-JUN-1993
; APPLICATION NUMBER: US 07/762682
; FILING DATE: 23-SEP-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/937195
; FILING DATE: 09-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Bonner, Grace L.
; REGISTRATION NUMBER: 32,963
; REFERENCE/DOCKET NUMBER: 38-21(10631)/A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (314)537-7286
; TELEFAX: (314)537-6047
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1647 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-393-785-13

```

Query Match	2.6%	Score 43.6;	DB 1;	Length 1647;	
Best Local Similarity	47.2%;	Prod. NO. 0.089;			
Matches 133;	Conservative	0;	Mismatches 149;	Indels 0;	Gaps 0;
QY	1109	ttttcaacacctaacttgggcactacttctgaagctgagcctgttggtctgggagcctgc	1168		
Db	971	TCTTCTCGGGCGCGCAGCCTCGGCTCCATTGAATGCTGTGCGCGCCGGGAGACCG	1030		
QY	1169	cggccctgatgtcatcgcgcactgggagtgccaaacagctacgcgcctctacaccatccgtg	1228		
Db	1031	GCACCTCTCCCGGGCTCAGTTCGAGATCGCGCGCGCTGGGGCCCCAACGGCAACATCA	1090		
QY	1229	ataggggaacgcacctctctgagctatgctgggttcctcgtagaaggacaaacatgtacg	1288		
Db	1091	TGACCGCCCGCGCCAAACATGTGTGGAACCCACGGCGCAGCAAGCAGTCGTGATCCCG	1150		
QY	1289	ccctctatatcacgtccacggtaactctctatcctaactctctcttggcatggtgtcc	1348		
Db	1151	CCCTCGGATCGACGATGGGACAAACCCCGACAAACCCCGCTCTGCGCGAGATAGCCCCA	1210		
QY	1349	tgccctggtggtggaagatcttcaccctgcccgtgcta	1390		
Db	1211	TGCCGGCGGGCTCGACACCTGGGTACGCTCTACCTGGCCA	1252		

RESULT 13
US-08-475-694-13
: Sequence 13, Application US/08475694
: Patent No. 5558862
: GENERAL INFORMATION:
: APPLICANT: Corbin, David R.
: APPLICANT: Greenplate, John T.
: APPLICANT: Jennings, Michael G.
: APPLICANT: Purcell, John P.
: APPLICANT: Sammons, Robert D.
: TITLE OF INVENTION: Method of Controlling Insects
: NUMBER OF SEQUENCES: 14
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: Grace L. Bonner, Monsanto Company, BB4F
: STREET: 700 Chesterfield Parkway No. 5558862th
: CITY: St. Louis

```

> STATE: Missouri
> COUNTRY: USA
> ZIP: 63198
> COMPUTER READABLE FORM:
> MEDIUM TYPE: Floppy disk
> COMPUTER: IBM PC compatible
> OPERATING SYSTEM: PC-DOS/MS-DOS
> SOFTWARE: PatentIn Release #1.0, Version #1.25
> CURRENT APPLICATION DATA:
> APPLICATION NUMBER: US/08/475,694
> FILING DATE: 07-JUN-1995
> CLASSIFICATION: 424
> PRIOR APPLICATION DATA:
> APPLICATION NUMBER: US 08/393,785
> FILING DATE: 24-FEB-1995
> APPLICATION NUMBER: US 08/083,948
> FILING DATE: 28-JUN-1993
> APPLICATION NUMBER: US 07/762682
> FILING DATE: 23-SEP-1991
> PRIOR APPLICATION DATA:
> APPLICATION NUMBER: US 07/937195
> FILING DATE: 09-SEP-1992
> ATTORNEY/AGENT INFORMATION:
> NAME: Bonner, Grace L.
> REGISTRATION NUMBER: 32,963
> REFERENCE/DOCKET NUMBER: 38-21(10631)A
> TELECOMMUNICATION INFORMATION:
> TELEPHONE: (314)537-7286
> TELEFAX: (314)537-6047
> INFORMATION FOR SEQ ID NO: 13:
> SEQUENCE CHARACTERISTICS:
> LENGTH: 1647 base pairs
> TYPE: nucleic acid
> STRANDEDNESS: single
> TOPOLOGY: linear
> MOLECULE TYPE: DNA (genomic)
> US-08-475-694-13

```

RESULT 14
US-08-712-057-13
Sequence 13, Application US/08712057
; Patent No. 5763245
; GENERAL INFORMATION:
; APPLICANT: Corbin, David R.
; APPLICANT: Greenplate, John T.
; APPLICANT: Purcell, John P.
; APPLICANT: Pershing, Jay

RESULT 15
US-08-083-948-7
; Sequence 7, Application US/08083948
; Patent No. 5518908
; GENERAL INFORMATION:
; APPLICANT: Corbin, David R.
; APPLICANT: Greenplate, John T.

	Query Match	2.6%	Score 43.6;	DB 1;	Length 1865;
	Best Local Similarity	47.2%;	Pred. No. 0.093;		
	Matches 133;	Conservative	0;	Mismatches 149;	Indels 0; Gaps 0;
Qy	1109	tcttcaacacctactctcggcactacttctctgaagctgagcctgtgtggtctggggcctgc	1168		
Db	1155	TCCTCTCGGCGCGCAGCCTCGGCTCCATTGAATGCTGCTGCGGCGCGGAGACCG	1214		
Qy	1169	cgcgcctgattgtcatcctggcactgggaatgccaacagcagtcgcctctacacatccgtg	1228		
Db	1215	GCACCTCTCGCGCGCTCAGTCCGAGATCGGCGCGGCTGGGGCCCCAACGGCAATCA	1274		
Qy	1229	ataggagaaacgcacactctctggagctatgtgttcctggaaggagacaaccatgtacg	1288		
Db	1275	TGACCGCGCGCGCAACCATGTGTGAAACCCACGCGGAGCAAGACGTCGTCGATCCCG	1334		
Qy	1289	ccctctatatcccggtccacggctacttccctcatcactctctcttctggcatggtgtcc	1348		
Db	1335	CCCTCGGCAATGACGACTGGGACAACCCCGACACCCCGTTCGCGGAGATAGCCCCCA	1394		
Qy	1349	tggccctgtgtgtggaagatcttccacctgtcccggtgcta	1390		
Db	1395	TGCGCGCGGCGCTCGAGACTGGGTGAGCTCTACTGGCCA	1436		

Search completed: June 11, 2002, 21:27:35
Job time: 4176 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 11, 2002, 21:26:13 ; Search time 61.01 Seconds
(without alignments)
999.500 Million cell updates/sec

Title: US-09-733-387-44

Perfect score: 2878

Sequence: 1 MATPRGLGALLLLLLLPTSG.....STTVSSSTARLDQASASQE 549

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 11073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_032802.*

1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2878	100.0	549	22 AAE04469	Human novel G-prot
2	2696	93.7	521	22 AAE04459	Human novel G-prot
3	2548	88.5	508	22 AAE04449	Human novel G-prot
4	2494	86.7	485	22 AAE04464	Human novel G-prot
5	2494	86.7	504	22 AAE04454	Human novel G-prot
6	2022	70.3	400	22 AAE04471	Human novel G-prot
7	1840	63.9	372	22 AAE04461	Human novel G-prot
8	1778	61.8	339	22 AAE04470	Human novel G-prot
9	1692	58.8	359	22 AAE04451	Human novel G-prot
10	1638	56.9	336	22 AAE04466	Human novel G-prot
11	1638	56.9	355	22 AAE04456	Human novel G-prot

12	1596	55.5	311	22 AAE04460	Human novel G-prot
13	1448	50.3	298	22 AAE04450	Human novel G-prot
14	1394	48.4	275	22 AAE04465	Human novel G-prot
15	1394	48.4	294	22 AAE04455	Human novel G-prot
16	1330	46.2	274	22 AAE04472	Human novel G-prot
17	1148	39.9	246	22 AAE04462	Human novel G-prot
18	1146	39.8	630	22 ABG14818	Novel human diagno
19	1094	38.0	225	22 AAE04448	Human novel G-prot
20	1089.5	37.9	585	22 ABH12463	Human bone marrow
21	1000	34.7	233	22 AAE04452	Human novel G-prot
22	946	32.9	210	22 AAE04467	Human novel G-prot
23	946	32.9	229	22 AAE04457	Human novel G-prot
24	896	31.1	211	22 ABB12380	Human bone marrow
25	804	27.9	157	22 AAE03245	Human gene 3 encod
26	799	27.8	158	22 AAE03244	Human gene 3 encod
27	671.5	23.3	203	22 AAE04473	Human novel G-prot
28	643.5	22.4	687	21 AAY79182	Haematopoietic ste
29	617	21.4	122	22 AAE03204	Human gene 3 encod
30	610.5	21.2	1013	21 AAB01247	Human HE6 receptor
31	610.5	21.2	1038	18 AAW36903	Human epididymis-s
32	607	21.1	128	22 AAE03246	Human gene 3 encod
33	603.5	21.0	512	22 AAE03246	Human BBSR seven t
34	603.5	21.0	687	21 AAB21700	Human TM7XN1. Hom
35	602.5	20.9	693	21 AAB01246	Human seven transm
36	599.5	20.8	687	20 AAY02382	Polypeptide identi
37	599.5	20.8	687	22 AAM93662	Human polypeptide,
38	596.5	20.7	693	20 AAY41765	Human PRO1083 prot
39	596.5	20.7	693	21 AAB44321	Human PRO1083 (UNQ
40	596.5	20.7	693	21 AAB33435	Human PRO1083 prot
41	596.5	20.7	693	21 AAY92362	G protein-coupled
42	596.5	20.7	693	22 AAM78369	Human protein SEQ
43	596.5	20.7	693	22 AAU12374	Human PRO1083 poly
44	594.5	20.7	693	20 AAY05730	Human G protein co
45	591.5	20.6	687	22 AAB88463	Human membrane or

ALIGNMENTS

RESULT 1

AAE04469
ID AAE04469 standard; Protein; 549 AA.

XX AAE04469;

DT 04-SEP-2001 (first entry)

XX Human novel G-protein coupled receptor (NGPCR) protein #22.

DE Human; novel G-protein coupled receptor; NGPCR; signal transduction;
KW drug screening; transmembrane protein; mental disorder; gene therapy;
KW nootropic; pharmacogenomics; behavioural disorder.

XX Homo sapiens.

XX WO200142287-A2.

XX 14-JUN-2001.

PF 07-DEC-2000; 2000WO-US33241.

XX 07-DEC-1999; 99US-0169427.

XX (LEXI-) LEXICON GENETICS INC.

PI Donoho G, Scooville J, Turner CA, Friedrich G, Zambrowicz B;

PI Sands AT;

XX WPI: 2001-381634/40.

XX N-PSDB: AAD08779.

PT Novel polynucleotides encoding human G protein coupled receptors useful
for drug screening, diagnosis and in gene therapy of physiological or

PT behavioral disorders -

PS Claim 2; Page 84-86; 91pp; English.

XX The present sequence is human novel G protein coupled receptor
XX (NGPCR) protein. NGPCRs are transmembrane proteins that span the
XX cellular membrane and are involved in signal transduction after ligand
XX binding. NGPCR DNA sequences are useful as hybridisation probes for
XX screening libraries and assessing gene expression patterns. Sequences
XX derived from regions adjacent to the intron/exon boundaries of NGPCR
XX gene can be used to design primers for use in amplification assays to
XX detect mutations within the exons, splice sites, introns that can be
XX used in diagnostics and pharmacogenomics. NGPCR sequences and/or gene
XX products are useful for identifying compounds that modulate gene
XX expression or gene product activity. Such compounds are useful in the
XX treatment of mental, physiological or behavioural disorders and diseases.
XX NGPCR DNA sequences and antibodies are useful for diagnostic and
XX prognostic evaluation of disorders related to NGPCR function and for the
XX identification of subjects having a predisposition to such disorders.
XX NGPCR DNA sequences are also useful for drug screening and in gene
XX therapy for modulating NGPCR expression and to produce genetically
XX engineered host cells to express NGPCR products in vivo. The encoded
XX NGPCR proteins are useful for generating antibodies, as reagents in
XX diagnostic assays and for identifying other cellular gene products
XX related to NGPCR.

XX Sequence 549 AA;

Query Match 100.0%; Score 2878; DB 22; Length 549;
Best Local Similarity 100.0%; Pred. No. 1.6e-289;
Matches 549; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MATPRGLGALLLLPTSGQEKTEGPRNTCLGSNNMYDIFNLNDKALCFTKCRQSGSD 60
Db 1 matprglgalllllptsgqektegrntclgsnnmydifnlndkalcftkcrqsgsd 60
Qy 61 SCNVENLQRYWLNVEAHLMEKGLTQKYNTPFLKALVONLSTNTAEDFYFSLPSQVPRQV 120
Db 61 scnvenlqrywlnveahlmekgltkyntpfalkalvnlstntaedfyfalepsqvrqv 120
Qy 121 MKDEKPPDRVRLPKSLFRSLPGNRSVVRVLAFTLIDPGTFLKPRGLGDSGVNLNR 180
Db 121 mkdekppdrvrlpkslfrslpgnrsvvravlftldpgtflkprglgdsgvnlmr 180
Qy 181 LVLSVGOMHVTKLAEPLEIVFSHQRPPTMTLCVFWDTKGTGDSSEGCSVEVRPE 240
Db 181 lvlsvgomhvtklaepileivfshqrpptmtlcvfwdvtkgtdgssegcsvevrpe 240
Y 241 GTVCCDHLTFFALLRPTLDQSTVHLITRISQAGCGVSMIFLAFTILLYAFILRSERF 300
Db 241 gtvccdhltffallrptldqstvhiltrisqagcgvsmlfaftillyafilrsrerf 300
Qy 301 KSEDAPKIHVALGSLFLNLAFVNVGSGKSDAACWARGAVFHYFLCAFTWGLEA 360
Db 301 ksedapkihalgsflnlafnlvngsgksdaacwargavfhyflcaftwmglea 360
Qy 361 FHLVLLAVRVNTVFHYFKLSLVGCLPALMVGITGSANSYGLKTIIRDRENTSLELC 420
Db 361 fhlyllavrvntvfhyfklslvvgclpalmvigtgsansyglytirdrenrtslelc 420
Qy 421 WFRGTTMYALYIVHGYFLITFLFGWVLLALVVWKITLSRATAVKERNKKVLTLL 480
Db 421 wfrgttmyalyivhgyflitflfgwvllalvvwkiltsratavkernkkrkvtll 480
Qy 481 GLSSLVGTWGLAFTPLGLSTVYIFALENSLQGVFCWFTTILYLPQSQTTSSTARL 540
Db 481 glsslvgtwglafthplglstvyifalenslogvfccwfttilylpqsqttvssstarl 540
Qy 541 DQAHSAQOE 549
Db 541 dqahsaqoe 549

RESULT 2

AAE04459
ID AAE04459 standard; Protein; 521 AA.

XX
AC AAE04459;

XX
DT 04-SEP-2001 (first entry)

XX
DE Human novel G-protein coupled receptor (NGPCR) protein #12.

XX
KW Human; novel G-protein coupled receptor; NGPCR; signal transduction;
KW drug screening; transmembrane protein; mental disorder; gene therapy;
KW nootropic; pharmacogenomics; behavioural disorder.

XX
OS Homo sapiens.

XX
PN WO200142287-A2.

XX
PD 14-JUN-2001.

XX
PF 07-DEC-2000; 2000WO-US33241.

XX
PR 07-DEC-1999; 99US-0169427.

XX
PA (LEXI-) LEXICON GENETICS INC.

XX
PI Donoho G, Scoville J, Turner CA, Friedrich G, Zambrowicz B;

XX
PI Sands AT;

XX
WPI; 2001-381634/40.

DR
N-PSDB; AAD08769.

XX
PT Novel polynucleotides encoding human G protein coupled receptors useful
PT for drug screening, diagnosis and in gene therapy of physiological or
PT behavioural disorders

PS
Disclosure: Page 73-74; 91pp; English.

XX The present sequence is human novel G protein coupled receptor
XX (NGPCR) protein. NGPCRs are transmembrane proteins that span the
XX cellular membrane and are involved in signal transduction after ligand
XX binding. NGPCR DNA sequences are useful as hybridisation probes for
XX screening libraries and assessing gene expression patterns. Sequences
XX derived from regions adjacent to the intron/exon boundaries of NGPCR
XX gene can be used to design primers for use in amplification assays to
XX detect mutations within the exons, splice sites, introns that can be
XX used in diagnostics and pharmacogenomics. NGPCR sequences and/or gene
XX products are useful for identifying compounds that modulate gene
XX expression or gene product activity. Such compounds are useful in the
XX treatment of mental, physiological or behavioural disorders and diseases.
XX NGPCR DNA sequences and antibodies are useful for diagnostic and
XX prognostic evaluation of disorders related to NGPCR function and for the
XX identification of subjects having a predisposition to such disorders.
XX NGPCR DNA sequences are also useful for drug screening and in gene
XX therapy for modulating NGPCR expression and to produce genetically
XX engineered host cells to express NGPCR products in vivo. The encoded
XX NGPCR proteins are useful for generating antibodies, as reagents in
XX diagnostic assays and for identifying other cellular gene products
XX related to NGPCR.

XX Sequence 521 AA;

Query Match 93.7%; Score 2696; DB 22; Length 521;
Best Local Similarity 100.0%; Pred. No. 1.1e-270;
Matches 514; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MATPRGLGALLLLPTSGQEKTEGPRNTCLGSNNMYDIFNLNDKALCFTKCRQSGSD 60

Db 1 matprglgalllllptsgqektegrntclgsnnmydifnlndkalcftkcrqsgsd 60

Qy 61 SCNVENLQRYWLNVEAHLMEKGLTQKYNTPFLKALVONLSTNTAEDFYFSLPSQVPRQV 120

||||| scvenlqrywlnyeahlmkegltqkntpfkalkvqnlstaeafyfslepgsqvprqv 120
Db 61 MKDEDKPPDRVRLPKSLFRSLPCNRSVVRVLAVTILDIGPCTLFKGPRLGLGDSGVLLNRR 180
Qy 121 MKDEDKPPDRVRLPKSLFRSLPCNRSVVRVLAVTILDIGPCTLFKGPRLGLGDSGVLLNRR 180
Db 121 MKDEDKPPDRVRLPKSLFRSLPCNRSVVRVLAVTILDIGPCTLFKGPRLGLGDSGVLLNRR 180
Qy 181 LVGLSVGQMHVTKLAEPLEIVFHSQRPNNMTLTCVFDVTKGTGDSSEGCSTEVRPE 240
Db 181 LVGLSVGQMHVTKLAEPLEIVFHSQRPNNMTLTCVFDVTKGTGDSSEGCSTEVRPE 240
Qy 241 GTVCCDHLTFFALLRPTLDQSTVHLTRISQAGCGVSMIFLAFTIILYAFRLSRERF 300
Db 241 GTVCCDHLTFFALLRPTLDQSTVHLTRISQAGCGVSMIFLAFTIILYAFRLSRERF 300
Qy 301 KSEDAPKIHVALGSLFLLNLAFVLNVGSGSGSDAACWARGAVHFYLLCAFTWMGLEA 360
Db 301 KSEDAPKIHVALGSLFLLNLAFVLNVGSGSGSDAACWARGAVHFYLLCAFTWMGLEA 360
Qy 361 FHYLLAVRVNTFYFGHYFLKLSLVGGLPALMVIGTGSANSYGLYTIIRDRENTSLELC 420
Db 361 FHYLLAVRVNTFYFGHYFLKLSLVGGLPALMVIGTGSANSYGLYTIIRDRENTSLELC 420
Qy 421 WREGTMYALYITVHGYYFLIFLFGMVVVALVVMKIFTLRSATAVKRGKRRKVVLTLL 480
Db 421 WREGTMYALYITVHGYYFLIFLFGMVVVALVVMKIFTLRSATAVKRGKRRKVVLTLL 480
Qy 481 GLSSLVGVTVWGLAIFTPGLSTVYIFALFNSLQG 514
Db 481 GLSSLVGVTVWGLAIFTPGLSTVYIFALFNSLQG 514

RESULT 3
AAE04449
ID AAE04449 standard; Protein; 508 AA.
XX AC AAE04449;
XX DT 04-SEP-2001 (first entry)
XX DE Human novel G-protein coupled receptor (NGPCR) protein #2.
XX KW Human; novel G-protein coupled receptor; NGPCR; signal transduction;
KW drug screening; transmembrane protein; mental disorder; gene therapy;
KW nootropic; pharmacogenomics; behavioural disorder.
XX OS Homo sapiens.
XX PN W0200142287-A2.
XX PD 14-JUN-2001.
XX PF 07-DEC-2000; 2000WO-US33241.
XX PR 07-DEC-1999; 99US-0169427.
XX PA (LEXI-) LEXICON GENETICS INC.
XX PI Donoho G, Scoville J, Turner CA, Friedrich G, Zambrowicz B;
PI Sands AT;
XX WPI; 2001-381634/40.
XX N-PSDB; AAD08759.
XX Novel polynucleotides encoding human G protein coupled receptors useful
PT for drug screening, diagnosis and in gene therapy of physiological or
PT behavioural disorders -
XX Claim 4; Page 62-63; 91pp; English.
XX The present sequence is human novel G protein coupled receptor
CC (NGPCR) protein. NGPCRs are transmembrane proteins that span the
CC cellular membrane and are involved in signal transduction after ligand

binding. NGPCR DNA sequences are useful as hybridisation probes for
screening libraries and assessing gene expression patterns. Sequences
derived from regions adjacent to the intron/exon boundaries of NGPCR
gene can be used to design primers for use in amplification assays to
detect mutations within the exons, splice sites, introns that can be
used in diagnostics and pharmacogenomics. NGPCR sequences and/or gene
products are useful for identifying compounds that modulate gene
expression or gene product activity. Such compounds are useful in the
treatment of mental, physiological or behavioural disorders and diseases.
NGPCR DNA sequences and antibodies are useful for diagnostic and
prognostic evaluation of disorders related to NGPCR function and for the
identification of subjects having a predisposition to such disorders.
NGPCR DNA sequences are also useful for drug screening and in gene
therapy for modulating NGPCR expression and to produce genetically
engineered host cells to express NGPCR products in vivo. The encoded
NGPCR proteins are useful for generating antibodies, as reagents in
diagnostic assays and for identifying other cellular gene products
related to NGPCR.

XX SQ Sequence 508 AA;

Query Match 88.5%; Score 2548; DB 22; Length 508;
Best Local Similarity 99.2%; Pred. No. 2.5e-255;
Matches 486; Conservative 1; Mismatches 1; Indels 2; Gaps 1;
Qy 1 MATPRGLGALLLLPTSGOEKPTGPRNCTCLGNNMNYDIFNLNDKALCFKCRQSGSD 60
Db 1 matprglgalllllptsggkptegprntclgnsnmmydfnlndkalcftkcrqsgsd 60
Qy 61 SCNVENLQRYWLNYEAHLMKEGLTKQVNTPELKVQLVQLNSTNTAEDEFYFSLEPSQVPRQV 120
Db 61 scnvenlqrywlnyeahlmkegltkqvntpeflkvalvqlnstntaedfyfslepsqvprqv 120
Qy 121 MKDEDKPPDRVRLPKSLFRSLPCNRSVVRVLAVTILDIGPCTLFKGPRLGLGDSGVLLNRR 180
Db 121 mkdedkppdrvrplpkslfrslpcnrsvvravlavtildigpctlfkgrpriqlgdsqvgvlnrr 180
Qy 181 LVGLSVGQMHVTKLAEPLEIVFHSQRPNNMTLTCVFDVTKGTGDSSEGCSTEVRPE 240
Db 181 lvglsvgqmhvtklaepleivfshqrpnnmtltcvfdvktgttdssegscstevrpe 240
Qy 241 GTVCCDHLTFFALLRPTLDQSTVHLTRISQAGCGVSMIFLAFTIILYAFRLSRERF 300
Db 241 gtvcccdhlthfallrptldqstvhilrtisqagcgysmiflaftiilyafrlsrerf 300
Qy 301 KSEDAPKIHVALGSLFLLNLAFVLNVGSGSGSDAACWARGAVHFYLLCAFTWMGLEA 360
Db 301 ksedapkihvalgslfllnlafvlvngsgsgsdacwargavfhyllcaftwmglea 360
Qy 361 FHYLLAVRVNTFYFGHYFLKLSLVGGLPALMVIGTGSANSYGLYTIIRDRENTSLELC 420
Db 361 fhyllylavrvntfyfghyflklslvgglpalmvigtgsansyglytirdrenrtslc 420
Qy 421 WREGTMYALYITVHGYYFLIFLFGMVVVALVVMKIFTLRSATAVKRGKRRKVVLTLL 480
Db 421 wregtmyalyitvhgyflitflfgmvvvalvvmkiftlrsatavkrgknrkvvltll 480
Qy 481 GLSSLVGVTVW 490
Db 481 glssl--asw 488

RESULT 4
AAE04464
ID AAE04464 standard; Protein; 485 AA.
XX AC AAE04464;
XX DT 04-SEP-2001 (first entry)
XX DE Human novel G-protein coupled receptor (NGPCR) protein #17.
XX

CC NGPCR proteins are useful for generating antibodies, as reagents in
CC diagnostic assays and for identifying other cellular gene products
CC related to NGPCR.
xx
SQ Sequence 504 AA:

DR WPI; 2001-381634/40.
DR N-PSDB; AAD08781.

DT 04-SEP-2001 (first entry)
 XX Human novel G-protein coupled receptor (NGPCR) protein #14.
 DE
 XX
 KW Human; novel G-protein coupled receptor; NGPCR; signal transduction;
 KW drug screening; transmembrane protein; mental disorder; gene therapy;
 KW nootropic; pharmacogenomics; behavioural disorder.
 XX
 OS Homo sapiens.
 XX
 XX WO200142287-A2.
 PN
 XX
 PD 14-JUN-2001.
 PF 07-DEC-2000; 2000WO-US33241.
 XX
 XX 07-DEC-1999; 99US-0169427.
 XX
 XX (LEXI-) LEXICON GENETICS INC.
 XX
 XX Donoho G, Scoville J, Turner CA, Friedrich G, Zambrowicz B;
 PI Sands AT;
 XX
 DR WPI; 2001-381634/40.
 DR N-PSDB; AAD08771.
 XX
 XX Novel polynucleotides encoding human G protein coupled receptors useful
 PT for drug screening, diagnosis and in gene therapy of physiological or
 PT behavioral disorders -
 XX
 PS Disclosure; Page 76-77; 91pp; English.
 XX
 CC The present sequence is human novel G protein coupled receptor
 CC (NGPCR) protein. NGPCRs are transmembrane proteins that span the
 CC cellular membrane and are involved in signal transduction after ligand
 CC binding. NGPCR DNA sequences are useful as hybridisation probes for
 CC screening libraries and assessing gene expression patterns. Sequences
 CC derived from regions adjacent to the intron/exon boundaries of NGPCR
 CC gene can be used to design primers for use in amplification assays to
 CC detect mutations within the exons, splice sites, introns that can be
 CC used in diagnostics and pharmacogenomics. NGPCR sequences and/or gene
 CC products are useful for identifying compounds that modulate gene
 CC expression or gene product activity. Such compounds are useful in the
 CC treatment of mental, physiological or behavioural disorders and diseases.
 CC NGPCR DNA sequences and antibodies are useful for diagnostic and
 CC prognostic evaluation of disorders related to NGPCR function and for the
 CC identification of subjects having a predisposition to such disorders.
 CC NGPCR DNA sequences are also useful for drug screening and in gene
 CC therapy for modulating NGPCR expression and to produce genetically
 CC engineered host cells to express NGPCR products in vivo. The encoded
 CC NGPCR proteins are useful for generating antibodies, as reagents in
 CC diagnostic assays and for identifying other cellular gene products
 CC related to NGPCR.
 XX
 XX Sequence 372 AA;
 SQ

Query Match 63.9%; Score 1840; DB 22; Length 372;
 Best Local Similarity 99.4%; Pred. No. 5.1e-182;
 Matches 351; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 162 LFKGPRLLGDSGVNRLNRLVGLSVGMVHTKLAEPLEIVFSHQRPPTNLTCVFDVDT 221
 | :|||||
 Db 13 lsgqprlgldgsgvlnrlvglsvgmhvklaepieivfshqrpptmtlcvfdvdt 72
 | :|||||
 Qy 222 KGTGDSWSECSFVRPEGVCCDHLTFALLRPTLDQSTVHILTRISQAGCGVSMI 281
 | :|||||
 Db 73 kgtgdsesecstevrpegvtcccdhlrtfllrptldqstvhiltrisgqcgvsml 132
 | :|||||
 Qy 282 FLATILYAFILRSRFRKSEDPKIHVALGSLFLNLAFILNVNVSQSGKSDAACWAR 341
 | :|||||
 Db 133 flatililyafilrsrfrksedapkhvalgslflnlafilnvngsgsgksdaacwar 192

Qy 342 GAVFHYELLCAFTWMLGELAFHLYLLAVRVNTYFVGHVFLKLSLVGWLPAVMVIGTGSAN 401
 | :|||||
 Db 193 gavfhyellcaftwmlgelaafhlyllavrvntfyghyflklslsvgwlpalvmvigtgsan 252
 | :|||||
 Qy 402 SYGLYIRDRNRTSLELCWFREGTMYALYIVVHGYFLITFLFGMVVIALVVKIFTLS 461
 | :|||||
 Db 253 syglyirdrentslelcwfregrtmyalyitvhgyflitflfgmvvialvwwkiftls 312
 | :|||||
 Qy 462 RATAVKERGKNRKKVLTLLGLSLVGVTVWGLAIFTPLGLSTVYVIFALFNSLOG 514
 | :|||||
 Db 313 ratavkergknrkkvtllglsslvgvtwglaitfplgltstvyvifalfnslsg 365
 | :|||||

RESULT 8
 AAE04470
 ID AAE04470 standard; Protein; 339 AA.
 XX
 AC AAE04470;
 XX
 DT 04-SEP-2001 (first entry)
 XX
 DE Human novel G-protein coupled receptor (NGPCR) protein #23.
 XX
 KW Human; novel G-protein coupled receptor; NGPCR; signal transduction;
 KW drug screening; transmembrane protein; mental disorder; gene therapy;
 KW nootropic; pharmacogenomics; behavioural disorder.
 XX
 OS Homo sapiens.
 XX
 XX WO200142287-A2.
 PN
 XX
 PD 14-JUN-2001.
 XX
 PF 07-DEC-2000; 2000WO-US33241.
 XX
 XX 07-DEC-1999; 99US-0169427.
 XX
 XX (LEXI-) LEXICON GENETICS INC.
 XX
 XX Donoho G, Scoville J, Turner CA, Friedrich G, Zambrowicz B;
 PI Sands AT;
 XX
 DR WPI; 2001-381634/40.
 DR N-PSDB; AAD08780.
 XX
 XX Novel polynucleotides encoding human G protein coupled receptors useful
 PT for drug screening, diagnosis and in gene therapy of physiological or
 PT behavioral disorders -
 XX
 PS Disclosure; Page 86-87; 91pp; English.
 XX
 CC The present sequence is human novel G protein coupled receptor
 CC (NGPCR) protein. NGPCRs are transmembrane proteins that span the
 CC cellular membrane and are involved in signal transduction after ligand
 CC binding. NGPCR DNA sequences are useful as hybridisation probes for
 CC screening libraries and assessing gene expression patterns. Sequences
 CC derived from regions adjacent to the intron/exon boundaries of NGPCR
 CC gene can be used to design primers for use in amplification assays to
 CC detect mutations within the exons, splice sites, introns that can be
 CC used in diagnostics and pharmacogenomics. NGPCR sequences and/or gene
 CC products are useful for identifying compounds that modulate gene
 CC expression or gene product activity. Such compounds are useful in the
 CC treatment of mental, physiological or behavioural disorders and diseases.
 CC NGPCR DNA sequences and antibodies are useful for diagnostic and
 CC prognostic evaluation of disorders related to NGPCR function and for the
 CC identification of subjects having a predisposition to such disorders.
 CC NGPCR DNA sequences are also useful for drug screening and in gene
 CC therapy for modulating NGPCR expression and to produce genetically
 CC engineered host cells to express NGPCR products in vivo. The encoded
 CC NGPCR proteins are useful for generating antibodies, as reagents in
 CC diagnostic assays and for identifying other cellular gene products
 CC related to NGPCR.
 XX

PR 07-DEC-1999; 99US-0169427.
XX (LEXI-) LEXICON GENETICS INC.
PA Donoho G, Scoville J, Turner CA, Friedrich G, Zambrowicz B;
PI Sands AT;
XX WPI: 2001-381634/40.
DR N-PSDB; AAD08776.
XX Novel polynucleotides encoding human G protein coupled receptors useful
PT for drug screening, diagnosis and in gene therapy of physiological or
PT behavioral disorders
XX Disclosure: Page 82; 91pp; English.
PS
XX The present sequence is human novel G protein coupled receptor
CC (NGPCR) protein. NGPCRs are transmembrane proteins that span the
CC cellular membrane and are involved in signal transduction after ligand
CC binding. NGPCR DNA sequences are useful as hybridisation probes for
CC screening libraries and assessing gene expression patterns. Sequences
CC derived from regions adjacent to the intron/exon boundaries of NGPCR
CC gene can be used to design primers for use in amplification assays to
CC detect mutations within the exons, splice sites, introns that can be
CC used in diagnostics and pharmacogenomics. NGPCR sequences and/or gene
CC products are useful for identifying compounds that modulate gene
CC expression or gene product activity. Such compounds are useful in the
CC treatment of mental, physiological or behavioural disorders and diseases.
CC NGPCR DNA sequences and antibodies are useful for diagnostic and
CC prognostic evaluation of disorders related to NGPCR function and for the
CC identification of subjects having a predisposition to such disorders.
CC NGPCR DNA sequences are also useful for drug screening and in gene
CC therapy for modulating NGPCR expression and to produce genetically
CC engineered host cells to express NGPCR products in vivo. The encoded
CC NGPCR proteins are useful for generating antibodies, as reagents in
CC diagnostic assays and for identifying other cellular gene products
CC related to NGPCR.
XX Sequence 336 AA;
SQ

Query Match 56.9%; Score 1638; DB 22; Length 336;
Best Local Similarity 99.4%; Pred. No. 4.1e-161;
Matches 310; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 162 LFKGPRGLGDSGVNLRNLVGLSVGQMHVTKLAEPLEIVFSHORPPNMTLTCVFWDVT 221
Db | : |||||
Qy 13 lsgqprlgldgsgvlnrlvlgsvgmhvtklaepleivfshqrpnmlltcvfdvt 72
Db |||||
Qy 222 KGTGDMSSGSGSTEVPEGTVCVCCDHLTFALLRPTLDQSTVHILTRISQAGCGVSMI 281
Db |||||
Qy 73 kgttdwssegcstevrpegtvcvccdhltffallrptldgstvhiltrisqagcgvsml 132
Db |||||
Qy 282 FLAFTIILYAFRLSRERFKSEDAPIKHVALGGSFLNLNLAFLNVNNGSGSGSDAACWAR 341
Db |||||
Qy 133 flaftiilayafirlsrerfkseadpkihvalggsflnlalnflnvngsgsgsdacwar 192
Qy 342 GAVFHYELLCRAFTWGLEAFLHYLLAVRVNTFYGHYFLKLSLVGGLPALMVIPTGSAN 401
Db |||||
Qy 193 gavihyflcaftwmgleathlyllavrvntfyghyflklsivgglpalmvigtgsan 252
Qy 402 SYGLTITRDENRTSLCLWFREGTWTYALYITVGHYFLITELFGMVVLALVVKFTLS 461
Db |||||
Qy 253 syglitirdenrtslclwfregtwtalyitvghyflitelfgmvvlaalvkvkftls 312
Qy 462 RATAVKERGNR 473
Db |||||
Qy 313 ratavkergnr 324

RESULT 11
AAE04456
ID AAE04456 standard; Protein; 355 AA.

XX AAE04456;
AC 04-SEP-2001 (first entry)
XX Human novel G-protein coupled receptor (NGPCR) protein #9.
DT
XX
DE
XX
XX Human; novel G-protein coupled receptor; NGPCR; signal transduction;
KW drug screening; transmembrane protein; mental disorder; gene therapy;
KW nootropic; pharmacogenomics; behavioural disorder.
XX
OS Homo sapiens.
XX WO200142287-A2.
PN
XX
XX 14-JUN-2001.
XX
XX 07-DEC-2000; 2000WO-US33241.
XX
XX 07-DEC-1999; 99US-0169427.
XX (LEXI-) LEXICON GENETICS INC.
XX
XX Donoho G, Scoville J, Turner CA, Friedrich G, Zambrowicz B;
PI Sands AT;
XX
XX WPI: 2001-381634/40.
DR N-PSDB; AAD08766.
XX Novel polynucleotides encoding human G protein coupled receptors useful
PT for drug screening, diagnosis and in gene therapy of physiological or
PT behavioural disorders
XX Disclosure: Page 70; 91pp; English.
PS
XX The present sequence is human novel G protein coupled receptor
CC (NGPCR) protein. NGPCRs are transmembrane proteins that span the
CC cellular membrane and are involved in signal transduction after ligand
CC binding. NGPCR DNA sequences are useful as hybridisation probes for
CC screening libraries and assessing gene expression patterns. Sequences
CC derived from regions adjacent to the intron/exon boundaries of NGPCR
CC gene can be used to design primers for use in amplification assays to
CC detect mutations within the exons, splice sites, introns that can be
CC used in diagnostics and pharmacogenomics. NGPCR sequences and/or gene
CC products are useful for identifying compounds that modulate gene
CC expression or gene product activity. Such compounds are useful in the
CC treatment of mental, physiological or behavioural disorders and diseases.
CC NGPCR DNA sequences and antibodies are useful for diagnostic and
CC prognostic evaluation of disorders related to NGPCR function and for the
CC identification of subjects having a predisposition to such disorders.
CC NGPCR DNA sequences are also useful for drug screening and in gene
CC therapy for modulating NGPCR expression and to produce genetically
CC engineered host cells to express NGPCR products in vivo. The encoded
CC NGPCR proteins are useful for generating antibodies, as reagents in
CC diagnostic assays and for identifying other cellular gene products
CC related to NGPCR.
XX Sequence 355 AA;
SQ

Query Match 56.9%; Score 1638; DB 22; Length 355;
Best Local Similarity 99.4%; Pred. No. 4.5e-161;
Matches 310; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 162 LFKGPRGLGDSGVNLRNLVGLSVGQMHVTKLAEPLEIVFSHORPPNMTLTCVFWDVT 221
Db | : |||||
Qy 13 lsgqprlgldgsgvlnrlvlgsvgmhvtklaepleivfshqrpnmlltcvfdvt 72
Db |||||
Qy 222 KGTGDMSSGSGSTEVPEGTVCVCCDHLTFALLRPTLDQSTVHILTRISQAGCGVSMI 281
Db |||||
Qy 73 kgttdwssegcstevrpegtvcvccdhltffallrptldgstvhiltrisqagcgvsml 132
Qy 282 FLAFTIILYAFRLSRERFKSEDAPIKHVALGGSFLNLNLAFLNVNNGSGSGSDAACWAR 341

Db 133 flatililyafirlsrrefksedapkihaigslflnlnlaflnvvgsgsksgdaacwar 192
QY 342 GAVFHYELLCFAFTWMLGAEAFHLYLLAVRVNTYFGHYFLKLSLVGWLGPALMVIIGTGSAN 401
Db 193 gavfhyflcaftwmgleafhlyllavrvntfyghyflklslvgwglpalmvigtgsan 252
QY 402 SYGLYTIRDRENRTSLELCWPREGTTMYALXIIVHGYFLITFLFGMWVIALVWVKIFTLS 461
Db 253 syglytirdrenrtslelcwrfregttmyalyitvhgyflitflfgmvvialvwwkiftls 312
QY 462 RATAVKRGKNR 473
Db 313 ratavkergknr 324
RESULT 12
AAE04460
ID AAE04460 standard; Protein; 311 AA.
XX AC AAE04460;
XX DT 04-SEP-2001 (first entry)
XX DE Human novel G-protein coupled receptor (NGPCR) protein #13.
XX KW Human; novel G-protein coupled receptor; NGPCR; signal transduction;
KW drug screening; transmembrane protein; mental disorder; gene therapy;
KW nootropic; pharmacogenomics; behavioural disorder.
XX OS Homo sapiens.
XX PN WO200142287-A2.
XX PD 14-JUN-2001.
XX PF 07-DEC-2000; 2000WO-US33241.
XX PR 07-DEC-1999; 99US-0169427.
XX PA (LEXI-) LEXICON GENETICS INC.
XX PI Donoho G, Scoville J, Turner CA, Friedrich G, Zambrowicz B;
XX Sands AT;
XX WPI: 2001-381634/40.
XX DR N-PSDB; AAD08770.
XX PT Novel polynucleotides encoding human G protein coupled receptors useful
XX for drug screening, diagnosis and in gene therapy of physiological or
XX behavioral disorders.
XX PS Disclosure; Page 75-76; 91pp; English.
XX CC The present sequence is human novel G protein coupled receptor
CC (NGPCR) protein. NGPCRs are transmembrane proteins that span the
CC cellular membrane and are involved in signal transduction after ligand
CC binding. NGPCR DNA sequences are useful as hybridisation probes for
CC screening libraries and assessing gene expression patterns. Sequences
CC derived from regions adjacent to the intron/exon boundaries of NGPCR
CC gene can be used to design primers for use in amplification assays to
CC detect mutations within the exons, splice sites, introns that can be
CC used in diagnostics and pharmacogenomics. NGPCR sequences and/or gene
CC products are useful for identifying compounds that modulate gene
CC expression or gene product activity. Such compounds are useful in the
CC treatment of mental, physiological or behavioural disorders and diseases.
CC NGPCR DNA sequences and antibodies are useful for diagnostic and
CC prognostic evaluation of disorders related to NGPCR function and for the
CC identification of subjects having a predisposition to such disorders.
CC NGPCR DNA sequences are also useful for drug screening and in gene
CC therapy for modulating NGPCR expression and to produce genetically
CC engineered host cells to express NGPCR products in vivo. The encoded
CC NGPCR proteins are useful for generating antibodies, as reagents in

CC diagnostic assays and for identifying other cellular gene products
CC related to NGPCR.
XX SQ Sequence 311 AA;
Query Match 55.5%; Score 1596; DB 22; Length 311;
Best Local Similarity 100.0%; Pred. No. 8.4e-157;
Matches 304; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 211 MLTCTCFWDTVKTGTDNSSEGCSTEVPRGTVCCDHLTFPALLRLPTLDQSTVHILTR 270
Db 1 mltctcfwdtvktgtdwssegcstevrptgtvcccdhltfallrlpldqstvhiltr 60
QY 271 ISOAGCGVSMIFLAFTIILYAFRLSRERFKSEDAPIKIHVALGGSLFLNLAFLVNVGSG 330
Db 61 isqagcgvsmiflaftiililyafirlsrerfksedapkihaigslflnlnlaflnvvgsg 120
QY 331 SKGSDAACWARGAVFHYFLLCFAFTWMLGAEAFHLYLLAVRVNTYFGHYFLKLSLVGWLGP 390
Db 121 skgsdaacwargavfhyflcaftwmgleafhlyllavrvntfyghyflklslvgwglp 180
QY 391 ALMVIIGTGSANSGLYTIRDRENRTSLELCWPREGTTMYALXIIVHGYFLITFLFGMWVL 450
Db 181 almvigtgsansglytirdrenrtslelcwrfregttmyalyitvhgyflitflfgmwvl 240
QY 451 ALVWVKIFTLSRATAVKERGNKRVLTILIGLSSLVGVTWGLAIFTPLGLSTVYIFALFN 510
Db 241 alvwvkiftlsratavkergknrkkvltlilgisslvgtwglaitfplglstvyifalfn 300
QY 511 SLOG 514
Db 301 slqg 304
RESULT 13
AAE04450
ID AAE04450 standard; Protein; 298 AA.
XX AC AAE04450;
XX DT 04-SEP-2001 (first entry)
XX DE Human novel G-protein coupled receptor (NGPCR) protein #3.
XX KW Human; novel G-protein coupled receptor; NGPCR; signal transduction;
KW drug screening; transmembrane protein; mental disorder; gene therapy;
KW nootropic; pharmacogenomics; behavioural disorder.
XX OS Homo sapiens.
XX PN WO200142287-A2.
XX PD 14-JUN-2001.
XX PF 07-DEC-2000; 2000WO-US33241.
XX PR 07-DEC-1999; 99US-0169427.
XX PA (LEXI-) LEXICON GENETICS INC.
XX PI Donoho G, Scoville J, Turner CA, Friedrich G, Zambrowicz B;
XX Sands AT;
XX WPI: 2001-381634/40.
XX DR N-PSDB; RAD08760.
XX PT Novel polynucleotides encoding human G protein coupled receptors useful
XX for drug screening, diagnosis and in gene therapy of physiological or
XX behavioral disorders.
XX PS Disclosure; Page 64; 91pp; English.
XX CC

CC The present sequence is human novel G protein coupled receptor
 CC (NGPCR) protein. NGPCRs are transmembrane proteins that span the
 CC cellular membrane and are involved in signal transduction after ligand
 CC binding. NGPCR DNA sequences are useful as hybridisation probes for
 CC screening libraries and assessing gene expression patterns. Sequences
 CC derived from regions adjacent to the intron/exon boundaries of NGPCR
 CC gene can be used to design primers for use in amplification assays to
 CC detect mutations within the exons, splice sites, introns that can be
 CC used in diagnostics and pharmacogenomics. NGPCR sequences and/or gene
 CC products are useful for identifying compounds that modulate gene
 CC expression or gene product activity. Such compounds are useful in the
 CC treatment of mental, physiological or behavioural disorders and diseases.
 CC NGPCR DNA sequences and antibodies are useful for diagnostic and
 CC prognostic evaluation of disorders related to NGPCR function and for the
 CC identification of subjects having a predisposition to such disorders.
 CC NGPCR DNA sequences are also useful for drug screening and in gene
 CC therapy for modulating NGPCR expression and to produce genetically
 CC engineered host cells to express NGPCR products in vivo. The encoded
 CC NGPCR proteins are useful for generating antibodies, as reagents in
 CC diagnostic assays and for identifying other cellular gene products
 CC related to NGPCR.

XX Sequence 298 AA;

Query Match 50.3%; Score 1448; DB 22; Length 298;
 Best Local Similarity 98.6%; Pred. No. 1.8e-141;
 Matches 276; Conservative 1; Mismatches 1; Indels 2; Gaps 1;

Qy 211 MTLTCVFDVDTGTTGDWSEGCSTVRPEGTVCDDHLLTFFALLRPTLDQSTVHLTR 270
 Db 1 mtltcvfdvdtgkttgdwsegcstevrpegtvcddhltfllrptldgstvhilr 60
 Qy 271 ISOAGCGVSMIFLAFTIILYAFRLSRERFKSEDPKIHVALGGSFLNLNLAFLVNGSG 330
 Db 61 isqagcgvmiflaftiilyafirlsrerfkseadpkihvalggsflnlaflnlvngsg 120
 Qy 331 SKGSDAACHWARGAVFHYFLLCAPTWMGLEAFHLYLLAVRVNTYFGHYFLKLSLVGNGLP 390
 Db 121 skgsdaacwargavfhyflilcawtmggleafhlyllavrvntfyghyflklslvngwlp 180
 Qy 391 ALMVIGTGSANSYGLYTIIRDRENTSLELCWFREGTMYALYITVHGYSFLITFLFGMYVL 450
 Db 181 almvigtgsansyglytirdrenrtselelcwfregtmyalyitvhgyltflfmgmyvl 240
 Qy 451 ALVWVKIFTLRSATAVKRGKKNR 490
 Db 241 alvwwkiftlrsatavkrgknrkvtlilgissl--asw 278

RESULT 14

AAE04465
 ID AAE04465 standard; Protein; 275 AA.
 XX
 AC AAE04465;

XX 04-SEP-2001 (first entry)

XX Human novel G-protein coupled receptor (NGPCR) protein #18.

XX Human; novel G-protein coupled receptor; NGPCR; signal transduction;
 XX drug screening; transmembrane protein; mental disorder; gene therapy;
 XX nootropic; pharmacogenomics; behavioural disorder.

XX Homo sapiens.

XX WO200142287-A2.

XX 14-JUN-2001.

XX 07-DEC-2000; 2000WO-US33241.

XX 07-DEC-1999; 99US-0169427.

XX (LEXI-) LEXICON GENETICS INC.
 PA Donoho G, Scoville J, Turner CA, Friedrich G, Zambrowicz B;
 PI Sands AT;
 XX WPI; 2001-381634/40.
 DR N-PSDB; AAD08775.

XX Novel polynucleotides encoding human G protein coupled receptors useful
 PT for drug screening, diagnosis and in gene therapy of physiological or
 PT behavioural disorders -

XX Disclosure; Page 81; 91pp; English.

XX The present sequence is human novel G protein coupled receptor
 CC (NGPCR) protein. NGPCRs are transmembrane proteins that span the
 CC cellular membrane and are involved in signal transduction after ligand
 CC binding. NGPCR DNA sequences are useful as hybridisation probes for
 CC screening libraries and assessing gene expression patterns. Sequences
 CC derived from regions adjacent to the intron/exon boundaries of NGPCR
 CC gene can be used to design primers for use in amplification assays to
 CC detect mutations within the exons, splice sites, introns that can be
 CC used in diagnostics and pharmacogenomics. NGPCR sequences and/or gene
 CC products are useful for identifying compounds that modulate gene
 CC expression or gene product activity. Such compounds are useful in the
 CC treatment of mental, physiological or behavioural disorders and diseases.
 CC NGPCR DNA sequences and antibodies are useful for diagnostic and
 CC prognostic evaluation of disorders related to NGPCR function and for the
 CC identification of subjects having a predisposition to such disorders.
 CC NGPCR DNA sequences are also useful for drug screening and in gene
 CC therapy for modulating NGPCR expression and to produce genetically
 CC engineered host cells to express NGPCR products in vivo. The encoded
 CC NGPCR proteins are useful for generating antibodies, as reagents in
 CC diagnostic assays and for identifying other cellular gene products
 CC related to NGPCR.

XX Sequence 275 AA;

Query Match 48.4%; Score 1394; DB 22; Length 275;
 Best Local Similarity 100.0%; Pred. No. 6.6e-136;
 Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 211 MTLTCVFDVDTGTTGDWSEGCSTVRPEGTVCDDHLLTFFALLRPTLDQSTVHLTR 270

Db 1 mtltcvfdvdtgkttgdwsegcstevrpegtvcddhltfllrptldgstvhilr 60

Qy 271 ISOAGCGVSMIFLAFTIILYAFRLSRERFKSEDPKIHVALGGSFLNLNLAFLVNGSG 330

Db 61 isqagcgvmiflaftiilyafirlsrerfkseadpkihvalggsflnlaflnlvngsg 120

Qy 331 SKGSDAACHWARGAVFHYFLLCAPTWMGLEAFHLYLLAVRVNTYFGHYFLKLSLVGNGLP 390

Db 121 skgsdaacwargavfhyflilcawtmggleafhlyllavrvntfyghyflklslvngwlp 180

Qy 391 ALMVIGTGSANSYGLYTIIRDRENTSLELCWFREGTMYALYITVHGYSFLITFLFGMYVL 450

Db 181 almvigtgsansyglytirdrenrtselelcwfregtmyalyitvhgyltflfmgmyvl 240

Qy 451 ALVWVKIFTLRSATAVKRGKKNR 473

Db 241 alvwwkiftlrsatavkrgknrkvtlilgissl--asw 263

RESULT 15

AAE04455
 ID AAE04455 standard; Protein; 294 AA.

XX
 AC AAE04455;

DT 04-SEP-2001 (first entry)

XX

DE Human novel G-protein coupled receptor (NGPCR) protein #8.
XX Human; novel G-protein coupled receptor; NGPCR; signal transduction;
KW drug screening; transmembrane protein; mental disorder; gene therapy;
KW nootropic; pharmacogenomics; behavioural disorder.
XX Homo sapiens.
OS
XX WO200142287-A2.
XX 14-JUN-2001.
XX 07-DEC-2000; 2000WO-US33241.
PF
XX 07-DEC-1999; 99US-0169427.
PR
XX (LEXI-) LEXICON GENETICS INC.
PA
XX Donoho G, Scoville J, Turner CA, Friedrich G, Zambrowicz B;
PI Sands AT;
PI
XX WPI: 2001-381634/40.
DR N-PSDB; AAD08765.
DR
XX Novel polynucleotides encoding human G protein coupled receptors useful
PT for drug screening, diagnosis and in gene therapy of physiological or
PT behavioral disorders
PT
XX Disclosure; Page 69-70; 91pp; English.
PS
XX The present sequence is human novel G protein coupled receptor
CC (NGPCR) protein. NGPCRs are transmembrane proteins that span the
CC cellular membrane and are involved in signal transduction after ligand
CC binding. NGPCR DNA sequences are useful as hybridisation probes for
CC screening libraries and assessing gene expression patterns. Sequences
CC derived from regions adjacent to the intron/exon boundaries of NGPCR
CC gene can be used to design primers for use in amplification assays to
CC detect mutations within the exons, splice sites, introns that can be
CC used in diagnostics and pharmacogenomics. NGPCR sequences and/or gene
CC products are useful for identifying compounds that modulate gene
CC expression or gene product activity. Such compounds are useful in the
CC treatment of mental, physiological or behavioural disorders and diseases.
CC NGPCR DNA sequences and antibodies are useful for diagnostic and
CC prognostic evaluation of disorders related to NGPCR function and for the
CC identification of subjects having a predisposition to such disorders.
CC NGPCR DNA sequences are also useful for drug screening and in gene
CC therapy for modulating NGPCR expression and to produce genetically
CC engineered host cells to express NGPCR products in vivo. The encoded
CC NGPCR proteins are useful for generating antibodies, as reagents in
CC diagnostic assays and for identifying other cellular gene products
CC related to NGPCR.
XX
SQ Sequence 294 AA;

Query Match 48.4%; Score 1394; DB 22; Length 294;
Best Local Similarity 100.0%; Pred. No. 7.3e-136;
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 MTLTCVFDVTKGTGWSSEGGSTEVPRPGTVCDDHLTFALLRLPTLDQSTVHLTR 270
DB 1 mtlctvfdvdkgtgdwssegcstevprpgtvcddhlrtfllrptldqstvhilrt 60
QY 271 ISOAGCGVSMFLAFTIILYAFLLSRERFKSEDAKIHVALGSLFLNLAFLVNMGSG 330
DB 61 isqagcgvsmflaftiilayfllsrerfkseadapkihvalggsflnlaflnvngs 120
QY 331 SKGSDAACWARGAVFHYFLLCFTWGLFAFLYLLAVRVNTYFGHYFLKLSLVGWLGP 390
DB 121 skgsdaacwargavfhyflcftwmgleafhlyllavrvfncyfghyflklslvgwglp 180
QY 391 ALMVIGTGSANSGLYTIIRDRENTSLELCWFREGTMTYALYITVHGYLELITFLFGMVYL 450
|||||

Db 181 almvigtgsansglytirdrenrtslelcwfregtmtyalYITVHGYLELITFLFGMVYL 240
QY 451 ALVVWKIFTLSRATAVKERKNR 473
Db 241 alvvwkiftlsratavkergknr 263

Search completed: June 11, 2002, 22:23:24
Job time: 3431 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 11, 2002, 21:31:23 ; Search time 24.88 seconds
(without alignments)
538.973 Million cell updates/sec

Title: US-09-733-387-44

Perfect score: 2878

Sequence: 1 MATPRGLGALLLLLLLPTSG.....STTVSSSTARLDQHASQSE 549

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA.*

1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*

2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*

3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*

4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*

5: /cgn2_6/ptodata/1/iaa/PTUS_COMB.pep.*

6: /cgn2_6/ptodata/1/iaa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	316	11.0	521	US-08-956-322-4	Sequence 4, Appli
2	316	11.0	652	US-09-110-116-1	Sequence 1, Appli
3	316	11.0	652	US-08-956-322-2	Sequence 2, Appli
4	300	10.4	344	US-09-110-116-4	Sequence 4, Appli
5	292.5	10.2	1052	US-08-852-806-2	Sequence 2, Appli
6	292.5	10.2	1052	US-09-163-669-2	Sequence 2, Appli
7	264	9.2	240	US-09-370-098-5	Sequence 5, Appli
8	254.5	8.8	886	US-09-110-116-3	Sequence 3, Appli
9	213	7.4	884	US-08-465-976A-2	Sequence 2, Appli
10	213	7.4	884	US-08-982-412-2	Sequence 2, Appli
11	199	6.9	235	US-09-370-098-6	Sequence 6, Appli
12	198.5	6.9	231	US-09-370-098-4	Sequence 4, Appli
13	177.5	6.2	415	US-08-110-286A-2	Sequence 2, Appli
14	177.5	6.2	431	US-08-981-189B-11	Sequence 11, Appli
15	174.5	6.1	431	US-08-381-433A-2	Sequence 2, Appli
16	173.5	6.0	411	US-08-381-433A-4	Sequence 4, Appli
17	173.5	6.0	411	US-08-981-189B-12	Sequence 12, Appli
18	173.5	6.0	431	US-08-981-189B-13	Sequence 13, Appli
19	172.5	6.0	415	US-08-110-286A-6	Sequence 6, Appli
20	172.5	6.0	415	US-08-981-189B-10	Sequence 10, Appli
21	158	5.5	509	US-08-845-546-2	Sequence 2, Appli
22	154.5	5.4	411	US-08-381-433A-8	Sequence 8, Appli
23	151	5.2	514	US-09-370-098-2	Sequence 2, Appli
24	149.5	5.2	553	US-08-845-546-12	Sequence 12, Appli
25	148.5	5.2	515	US-08-468-249A-18	Sequence 18, Appli
26	148.5	5.2	585	US-08-142-439A-6	Sequence 6, Appli
27	148.5	5.2	585	US-08-142-551B-125	Sequence 125, App

28 148.5 5.2 585 2 US-08-869-477-6 Sequence 6, Appli
29 148.5 5.2 585 2 US-08-468-249A-19 Sequence 19, Appli
30 143 5.0 482 1 US-08-142-439A-7 Sequence 7, Appli
31 143 5.0 482 2 US-08-869-477-7 Sequence 7, Appli
32 143 5.0 541 3 US-08-468-011A-2 Sequence 2, Appli
33 143 5.0 541 4 US-09-236-468A-2 Sequence 2, Appli
34 143 5.0 541 5 PCT-US95-07085-2 Sequence 2, Appli
35 140 4.9 482 1 US-07-792-885A-1 Sequence 1, Appli
36 140 4.9 593 2 US-08-468-249A-21 Sequence 21, Appli
37 138 4.8 269 4 US-09-370-098-3 Sequence 3, Appli
38 134 4.7 437 2 US-08-538-816A-2 Sequence 2, Appli
39 134 4.7 437 2 US-09-076-651-2 Sequence 2, Appli
40 134 4.7 437 4 US-09-208-394-2 Sequence 2, Appli
41 133.5 4.6 472 5 PCT-US94-09235-2 Sequence 2, Appli
42 133.5 4.6 474 1 US-08-453-742-2 Sequence 2, Appli
43 133.5 4.6 474 1 US-08-454-464-2 Sequence 2, Appli
44 133.5 4.6 474 1 US-08-453-222-2 Sequence 2, Appli
45 133.5 4.6 474 1 US-08-452-802-2 Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-956-322-4
; Sequence 4, Application US/08956322
; Patent No. 6277977
; GENERAL INFORMATION:
; APPLICANT: SATHE, GANESH
; APPLICANT: MAO, JOYCE
; TITLE OF INVENTION: CDNA CLONE HAPO167 THAT ENCODES
; TITLE OF INVENTION: A HUMAN 7-TRANSMEMBRANE RECEPTOR
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: RATNER & PRESTIA
; STREET: P.O. BOX 980
; CITY: VALLEY FORGE
; STATE: PA
; COUNTRY: USA
; ZIP: 19482
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/956,322
; FILING DATE: 23-OCT-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/049,329
; FILING DATE: 11-JUN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: PRESTIA, PAUL F
; REGISTRATION NUMBER: 23,031
; REFERENCE/DOCKET NUMBER: GH-70075
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-407-0700
; TELEFAX: 610-407-0701
; TELEX: 846169
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 521 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-956-322-4

Query Match 11.0%; Score 316; DB 4; Length 521;
Best Local Similarity 28.6%; Pred. No. 1.4e-23;
Matches 105; Conservative 55; Mismatches 165; Indels 42; Gaps 16;

QY 295 LSREKESDAPKIHVALGSLFLNLAFVLNVGSGKSDAACWARGAVHYFLICAPT 354
Db 383 AIONTS-----LHLQSLCLFLAHLFL--VGIDRTEPKVLCSSIIAGALHYLYAAFT 435
QY 355 WMGLEAFHYLYA--VRVENTYFGHVFLKLSL--VWGLPALMVIGTGSANSYGLYTRD 410
Db 436 WMLLEGVHLFLTARNLTVYVYSSINRLMKWIMFPVGVGVA--VTVAISAASWPHLYGTAD 494
QY 411 RENRTSLELCWREGTMYALYTVHGYFLITFLGMVVALVWV----KIFTLSTRATV 466
Db 495 R-----CWLHLOGFMWSEL---GPVCAIFSANLVFLFVILKRLKLSLSEYST 543
QY 467 KERKNNKVKVLTLLGLSSL--VGVTW--GLAIFTPLGLSTVYIFALENSLOGVFCWFT 522
Db 544 IQ-----NTRMLAFKATAQLFILGCTWCLGLLQVGAQVMAVLFYIINSLQGGFI--FL 596
QY 523 ILYLPQS 529
Db 597 VYCLLSQ 603

RESULT 4
US-09-110-116-4
; Sequence 4, Application US/09110116
; Patent No. 6013479
; GENERAL INFORMATION:
; APPLICANT: Xu, Hong
; APPLICANT: Cohan, Victoria L.
; APPLICANT: Stuart, Susan G.
; TITLE OF INVENTION: HUMAN EMRL-LIKE G PROTEIN COUPLED
; TITLE OF INVENTION: RECEPTOR
; FILE REFERENCE: PF-0550 US
; CURRENT APPLICATION NUMBER: US/09/110,116
; CURRENT FILING DATE: 1998-07-02
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 344
; TYPE: PRT
; ORGANISM: HOMO SAPIENS
; FEATURE:
; OTHER INFORMATION: 2935597, GenBank
US-09-110-116-4

Query Match 10.4%; Score 300; DB 3; Length 344;
Best Local Similarity 29.4%; Pred. No. 3.2e-22;
Matches 95; Conservative 50; Mismatches 148; Indels 30; Gaps 12;

QY 215 CVFWDVTGTTGDWSSEGST-EVRPEGTVCCHLTFALLRLPTLDQSTVHILTRISQ 273
Db 4 CVFWEHGQCGCHWATTCSTIGTRDTSTICRCTLHLSFAVLM-AHYDVQEDPVLTVITY 62
QY 274 ACGGSMIFLAFTIILYAFRLSRERKSEDAKPIHVALGSLFLNLAFVLNVGSGSKG 333
Db 63 MGLSVSLCLLLAATFLCKAIQNTS-----LHLQSLCLFLAHLFLVAT--DOTG 115
QY 334 SDAACWARGAVHYFLLCFAFTWGLAFHYLYLA--VRVENTYFGHVFLKLSL--VWGL 389
Db 116 HVLCSIIAGTHLYLYLATFTWMLLEALYLFETARNLTVYVYSSINRPMKMLFPVGV 175
QY 390 PALMVGITGSANSYGLYTRDRENTSLELCWFR-EGTMYALYTVHGYFLITFLGMV 448
Db 176 PA-VTVAISAASRPHLYTPSR-----CWLQPEKGFINGFLGPVCAIFSNVLVFLV 226
QY 449 VLAIVVWKIFTLSTRATYKRGKNNKVKVLTLLGLSSLGVTVWGLAIFT--PLGLSTVIIF 506
Db 227 TLWLKLNKLSLN--SEVSTLRNTRMLAFKATAQLFILGCTWCLGLILQVGAARVMAYLF 284
QY 507 ALFNSLOGVFCWFTIYLPQS 529
Db 285 TIINSLOGVFI--FLVYCLLSQ 304

RESULT 5
US-08-852-806-2
; Sequence 2, Application US/08852806
; Patent No. 5874245
; GENERAL INFORMATION:
; APPLICANT: Shoji Fukusumi
; APPLICANT: Shuji Hinuma
; TITLE OF INVENTION: NOVEL HUMAN G-PROTEIN COUPLED RECEPTOR
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: U.S.A.
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/852,806
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/017,915
; FILING DATE: 16 MAY 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Han, William T
; REGISTRATION NUMBER: 34,344
; REFERENCE/DOCKET NUMBER: TAK50002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5219
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1052 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
US-08-852-806-2

Query Match 10.2%; Score 292.5; DB 2; Length 1052;
Best Local Similarity 28.2%; Pred. No. 1.1e-20;
Matches 127; Conservative 57; Mismatches 186; Indels 81; Gaps 23;

QY 118 RQVMKDBDK-PDDRVRLPKS-LFRSLPGRNSVVRVLAFTI-LDIGPGLTKFGP-----RLG 169
Db 245 RRGKDWVRSEDRLELPKEVLSLSPGKPAATSGAGSGRGRGPGTVPVPGHSHQRL 304
QY 170 LGDGS-----GVLANRLVGLSVGMHVTKLAEPLEIVFHSORPP-----PNMFLT-- 214
Db 305 PADPDESSYEVIGAVLYRTGLILPPPR-PPLAVTSRVMTVTVPPTQPPAEPLIIVLS 363
QY 215 -----CVFWDYTK--GTTGDSWSEGST-EVRPEGTVCCHLTFALLRLPTLD 261
Db 364 YIINGTDPHCASWYSRADSSDQDWTENCQTLTAAHTRCQCQHLSTFAVLAQPPK 423
QY 262 QSTVHILTRIS---QAGCGVSMIFLAFTIILY-AFRLSRERKSEDAKPIHVALGSLF 317
Db 424 -LTLELAGSPVPLVIGCAVSCMALITLAIYAAFWRF-----IKSE---RSILNFCUS 475
QY 318 LNLAFVLNVGSGKSDAACWARGAVHYFLLCFAFTWGLAFHYLYLLAV-RVENTYFG 376

Db 476 ILASNILIVGOSRVLSKGVCTMTAAFLHFFFLSSFCVLTAEQSYLAVIGRMTRLVR 535
QY 377 HYFLKLSLVGGLPALMV---IGTSANSYGLYTIIRDRENTSLELCWFR-BGTTMYALY 432
Db 536 KRFLCL--GWGLPALVAVSVGFTRTKYG-----TSSYCWLSLEGGLLYAFV 581
QY 433 ITVHGVELITFLFGMVVLAVVWKIFTLRSATVAVKRGKNEKKVLTLLG---LSSLVGV 489
Db 582 GPAAVIVLNMGLIIVFN-----KLMARDGISDKSKKORAGASLWSSCVLPLLAIT 634
QY 490 WGLAIFTPLGLSTVY---IFALFNSLOGVFI 517
Db 635 WMSAVLMTDRRSVLFQALFAVNSAQGFVI 665
RESULT 6
US-09-163-669-2
; Sequence 2, Application US/09163669
; Patent No. 611076
; GENERAL INFORMATION:
; APPLICANT: FUKUSUMI, SHOJI
; APPLICANT: HINUMA, SHUJI
; APPLICANT: FUJII, RYO
; TITLE OF INVENTION: NOVEL HUMAN G-PROTEIN COUPLED
; TITLE OF INVENTION: RECEPTOR (HIBCD07)
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ratner & Prestia
; STREET: P.O. Box 980
; CITY: Valley Forge
; STATE: PA
; COUNTRY: USA
; ZIP: 19482
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA: US/09/163,669
; FILING DATE: 30-SEP-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/852,806
; FILING DATE: 07-MAY-1997
; APPLICATION NUMBER: 60/017,915
; FILING DATE: 16-MAY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Prestia, Paul F
; REGISTRATION NUMBER: 23,031
; REFERENCE/DOCKET NUMBER: TAK-50002-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-407-0700
; TELEFAX: 610-407-0700
; TELEX: 846169
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1052 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-163-669-2
Query Match 10.2%; Score 292.5; DB 3; Length 1052;
Best Local Similarity 28.2%; Pred. No. 1.1e-20;
Matches 127; Conservative 57; Mismatches 186; Indels 81; Gaps 23;
QY 118 ROYMKDEK-PPDRVRLPKS-LFRSLPGNRSVVRLAVTI-LDIGPTLFKGP-----RLG 169
Db 245 RRGKMDWRHSEDRLFLPKFVLSLSPGKPKATSGAUGPGRGPGTVPDPGSHORLL 304

QY 170 LGDGS-----CVLNNRLVGLSVGQMHVTKLAEPLEIVFSHORPB-----PNMTLT-- 214
Db 305 PADDESSYFVIGAVLYRTGLLPPPR-PPLAVTSRVMTVTRPTQPPAEPLEIIVELS 363
QY 215 -----CVFWDVK--GTGDMSSGECST-EVRPEGTVCCDHLFFALLRPTLD 261
Db 364 YIINGTTDPHCASWDYSRADASSGDWDTCQLETAHAHTRCQCHLSTFAVLAOPPKD 423
QY 262 QSTVHLITRIS---QAGCGVSMIFLAFTIILY-AFURLSRERFKSDEAPKIHVALGGSIF 317
Db 424 -LTLELAGSPVPLVICCAVSCMALTLAIYAAFWRF---IKSE---RSIILLNFCLS 475
QY 318 LLNLAEFLVNVGSGKSDAACWARGAVFHYFLLCATWMLGAEAFHLYLLAV-RVNTYFG 376
Db 476 ILASNILIVGQSRVLSKGVCTMTAAFLHFFFLSSFCVLTAEQSYLAVIGRMTRLVR 535
QY 377 HYFLKLSLVGGLPALMV---IGTSANSYGLYTIIRDRENTSLELCWFR-BGTTMYALY 432
Db 536 KRFLCL--GWGLPALVAVSVGFTRTKYG-----TSSYCWLSLEGGLLYAFV 581
QY 433 ITVHGVELITFLFGMVVLAVVWKIFTLRSATVAVKRGKNEKKVLTLLG---LSSLVGV 489
Db 582 GPAAVIVLNMGLIIVFN-----KLMARDGISDKSKKORAGASLWSSCVLPLLAIT 634
QY 490 WGLAIFTPLGLSTVY---IFALFNSLOGVFI 517
Db 635 WMSAVLMTDRRSVLFQALFAVNSAQGFVI 665
RESULT 7
US-09-370-098-5
; Sequence 5, Application US/09370098
; Patent No. 6303768
; GENERAL INFORMATION:
; APPLICANT: Lin, Yi-Jyun
; APPLICANT: Benzer, Seymour
; APPLICANT: California Institute of Technology
; TITLE OF INVENTION: METHUSELAH GENE, COMPOSITIONS AND
; TITLE OF INVENTION: METHODS OF USE
; FILE REFERENCE: 06618/343001
; CURRENT APPLICATION NUMBER: US/09/370,098
; PRIOR FILING DATE: 1999-08-06
; PRIOR APPLICATION NUMBER: 60/095,826
; PRIOR FILING DATE: 1998-08-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 240
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-09-370-098-5
Query Match 9.2%; Score 264; DB 4; Length 240;
Best Local Similarity 31.5%; Pred. No. 8.7e-19;
Matches 84; Conservative 37; Mismatches 98; Indels 48; Gaps 11;

QY 267 ILTRISQAGCGVSMIFLAFTIILYAFRLSRERFKSDEAPKIHVALGGSIFLLNLAFV 326
Db 4 LLSVITWVGIVISLVCLAIICISTFCFL-----RGLQTDRTTHKNLCINLFLAELLFL-- 56
QY 327 VSGSGKSDAACWARGAVFHYFLLCATWMLGAEAFHLYLLAVRVNTYFG---HYFLKLS 383
Db 57 VGIDKQYEVACPIFAGLLHYFFFLAAFSWLCLEGVHLYLLLVFVFESEYSRKYYILG-- 114
QY 384 LVWGGLPALMVIGTGSANSYGLYTIIRDRENTSLELCWFRGTTMYALYITVHGVELITF 443
Db 115 --GYCFPAL-VWGIAAAIDRYSYGT-----KACWLR-----VDNYFINWF 152
QY 444 LFGMVVLAVVWKF-----TLSRATVAVKRGKNE-----KKVLTLLGLSSLVGVW-- 490
Db 153 I-GPVSEVIVVNLVFLMVLHKMIRSSSVLKDPSRLDNKSWALCAIALLLGLGTWAF 211

QY 491 GLAIFTPLGLSTVYIFALFNSLQGVFI 517
Db 212 GLLFINKESVWYAYLFTTFNAFGGVFI 238

RESULT 8
US-09-110-116-3
; Sequence 3, Application US/09110116
; Patent No. 6013479
; GENERAL INFORMATION:
; APPLICANT: Xu, Hong
; APPLICANT: Cohan, Victoria L.
; APPLICANT: Stuart, Susan G.
; TITLE OF INVENTION: HUMAN EMRI-LIKE G PROTEIN COUPLED
; FILE REFERENCE: PF-0550 US
; CURRENT APPLICATION NUMBER: US/09/110,116
; CURRENT FILING DATE: 1998-07-02
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 886
; TYPE: PRT
; ORGANISM: HOMO SAPIENS
; FEATURE:
; OTHER INFORMATION: 784994, GenBank
US-09-110-116-3

Query Match 8.8%; Score 254.5; DB 3; Length 886;
Best Local Similarity 23.2%; Pred. No. 6.1e-17;
Matches 155; Conservative 84; Mismatches 232; Indels 197; Gaps 39;

QY 21 QEPTEGPRNTCLGSNNMYDI-----FNLNDKALCFKCRQSGSDSCNVENLORYWL---- 72
Db 274 QDPSTCGPNSICNALGSLVSCGIVGFHPNPEG-----SOKDGNFSC-----QRLVFKCK 323

QY 73 -----NYEAHLMKEGTLQK-----VNTPF--LKALVQNLST-----NTAEDFYFS 110
Db 324 EDVIPDNKQIQOCEGTAVKPAVVSFCAQINNIFSLVDKVCENKTTIVSLKNTTESFVPV 383

QY 111 LEPSQVPRQVMKDED-----KPPDRVRLPKSLFSLPGRNRSVRL 150
Db 384 LKOISMTWTKFTKETSLSATVLFVESVMTLASFWKPSANV-TPAVRAEYLDIESKVINK 442

QY 151 AVTILDIGPGLKPGRLGLG-----DCSGV----- 176
Db 443 ECSEENVTLDLVAKGDKMKIGCSTIESESTETTGVAFSVGMESVLNERFFQDHOAPL 502

QY 177 -----LNNRLVGLSVGMHYTK---LAEPLEIVFHSQRPNNMTL---TCVFDV-T 221
Db 503 TTSEIKLKMSRVVG---GIMTGEKKDGFSDP---IIVTLENVQPKQKFERPICVSWSTDV 557

QY 222 KGTGDMSSGCG-STEVRPBGTVCCDHLTFPALLRP---TLDSQTVHILTRISQAGCG 277
Db 558 KG--GRWTSFCVILEASRTYITCSCNQMANLAVIMASGELTMDFS-----LYIISHVGII 611

QY 278 VSMIFLAFTIILAFRLSRERKSEDAKPIHVALGSGSLFLNLAFILVNVGSGKSGDAA 337
Db 612 ISLVCLVLAATATLLCSIRNH-----NTYLHLHLVCVLLLAATLFL--AGIHKTONTKG 664

QY 338 CWARGAVFHYFLLCFAFTWGLFAFLYLLA--VRVFNTEYFGHYFLKLSLV---GWLGPAL 392
Db 665 CAITAGFLHYFLACFFWMLVEAVILFLMVRLKVVN-YFSSRNKMLHICAFYGLPML 723

QY 393 MVIOTGSA--NSYGLXYTRDRENTSLELCWFREGTMYALYITVHGYYFLITFLFGMVL 450
Db 724 VVVISASVQOGYGMH-----NR-----CWLNTETG-----FTWSEL-GPVCT 760

QY 451 ALVV-----WKITFL-SRATAKERKNRK--KVLTLGLSL--VGVTWGLAIFT--P 497
Db 761 VIVINSLLLTWTLWRQLRSSNAEYSTLTKDTRLLTFKFAQLFILGCSWILGIFQIGP 820

QY 498 LGLSTVYIFALFNSLQGVFI-----CC-----WFTILYLPQSSTTVSSSTAR--L 540
Db 821 VAGWYAVLFTIINSLOQAFILHCLLNGQVREYKRWITKTKPSS-----QSOTSRIILL 876

QY 541 DOAHSASQ 548
Db 877 SSMPASQ 884

RESULT 9
US-08-465-976A-2
; Sequence 2, Application US/08465976A
; Patent No. 5869632
; GENERAL INFORMATION:
; APPLICANT: SOPPET, DANIEL R
; APPLICANT: LI, YI
; APPLICANT: ROSEN, CRAIG A
; APPLICANT: RUBEN, STEVEN M
; TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELIA, BYRNE, BAIN GILFILLAN, CECCHI
; ADDRESSEE: STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NJ
; COUNTRY: US
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,976A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: FERRARO, GREGORY F
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 325800-444
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (201) 994-1700
; TELEFAX: (201) 994-1744
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 884 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-465-976A-2

Query Match 7.4%; Score 213; DB 2; Length 884;
Best Local Similarity 24.4%; Pred. No. 1e-12;
Matches 96; Conservative 61; Mismatches 163; Indels 74; Gaps 19;

QY 177 LNNRLVGLSV--GOMHVTKLAE-PLEIVFHSQRPNNMTLTCTVFDVDTKGT--GDWSS 230
Db 1 MNSPVSVAVFGRFLRGILESPISLFRLLQTNARSKAICVQND-PGLAEQHGVTWA 59

QY 231 EGCSTEVRPEGT--VCCDHLTFPALLL-----RPTLDOST-----VHILTRISQAGCGV 278
Db 60 RDECL-VHRNGSHARCRCSTCTGFGVLMDSAPRERLEGDLELLAVFTHVAVSVA---- 114

QY 279 SMIFLAFTIILAFRLSRERKSEDAKPIHVALGSGSLFLNLAFILVNVGSGKSGDAA 338
Db 115 -----ALVFAAIIILSLRSLS-KNVRGIHANVAALGVAELFL--LGIHRTNQLVC 164

QY 339 WARGAVFHYFLLCFAFTWGLFAFLYLLAVRVFNT-----YFGHYFLKLSLVGWLGPAL 393
Db 165 TAVAILLHYFLSTFAWLFVQGLHLYRMOVEPRNDRGNMRFYH-----ALGMGVPAVL 218

QY 394 VIGTGSANSYGLYTIIRDRENTSLELCWFREGTMYALYITVHGXYFLITFLFGMVVLALV 453
Db 219 L-----GLAVGLDPEGYGNPDFCW-----ISVHEPLIWSFA-GPVVLVIV 257
QY 454 V-WKIFTLSRATAVKRGKRNKVVLTLLGLSS-----LVGVTW--GLAIFTPLGLSTVYI 505
Db 258 MNGTWFLAARTSCSTGQREAKKTSALTIRSFLLLLLVLSASWLFGLLAVNHSILAFHYL 317
QY 506 FALFNSLQGVFICCFWFTIILYLPQSQTIVSSSTAR 539
Db 318 HAGLGLQGLAVLLFCVLNADARAAMPACIGR 351
RESULT 10
US-08-982-412-2
; Sequence 2, Application US/08982412
; Patent No. 5958729
; GENERAL INFORMATION:
; APPLICANT: SOPPET, DANIEL R
; APPLICANT: LI, YI
; APPLICANT: ROSEN, CRAIG A
; APPLICANT: RUBEN, STEVEN M
; TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HUMAN GENOME SCIENCES, INC.
; STREET: 9410 KEY WEST AVENUE
; CITY: ROCKVILLE,
; STATE: MD
; COUNTRY: US
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/982,412
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: BROOKES, ANDERS A
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PF161PCT2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8439
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 884 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-982-412-2

Query Match 7.4%; Score 213; DB 2; Length 884;
Best Local Similarity 24.4%; Pred. No. 1e-12;
Matches 96; Conservative 61; Mismatches 163; Indels 74; Gaps 19;
QY 177 LNNRLVGLSV--GQMHYTKLAE-PLEIVFSHQRPNNMTLTCVFDVTKGTT---GDWSS 230
Db 1 MNSPVSVAVFHGRNFRGLGILESPISLEFRLLQTNANRKAICQWQD--PPGLAEQHGHWTA 59
QY 231 EGCSTVRPGT--VCCCDHLTFALL-----RPTLDQST-----VHILTRISQACGV 278
Db 60 RDCEL-VHRNGSHARCRCRTGTGTGVLMDASPRERLEGDLLELAVFTHVVAVSVA--- 114
QY 279 SMIFLAFTIILYAFRLRSRERFKSEDAKPIHVALGSLFLNLAFLVNVGSGKSDAAC 338
Db 115 -----ALVLTAILLSLSLSKLS-NVRGIHANVAALGAELLLFL--LGIHRTHNQLVC 164
QY 339 WARGAVFHYLLCAFTWMGLEAFHYLLAVRVNT-----YFGHYELKLSLVGWGLPALM 393

Db 165 TAVAILLHYFFLSFAWLFVQGLHLYRMQVEPRNVDRGAMRFYH-----ALGCGVPAVL 218
QY 394 VIGTGSANSYGLYTIIRDRENTSLELCWFREGTMYALYITVHGXYFLITFLFGMVVLALV 453
Db 219 L-----GLAVGLDPEGYGNPDFCW-----ISVHEPLIWSFA-GPVVLVIV 257
QY 454 V-WKIFTLSRATAVKRGKRNKVVLTLLGLSS-----LVGVTW--GLAIFTPLGLSTVYI 505
Db 258 MNGTWFLAARTSCSTGQREAKKTSALTIRSFLLLLLVLSASWLFGLLAVNHSILAFHYL 317
QY 506 FALFNSLQGVFICCFWFTIILYLPQSQTIVSSSTAR 539
Db 318 HAGLGLQGLAVLLFCVLNADARAAMPACIGR 351
RESULT 11
US-09-370-098-6
; Sequence 6, Application US/09370098
; Patent No. 6303768
; GENERAL INFORMATION:
; APPLICANT: Benzer, Seymour
; APPLICANT: Lin, Yi-Jyun
; APPLICANT: California Institute of Technology
; TITLE OF INVENTION: METHUSELAH GENE, COMPOSITIONS AND
; FILE REFERENCE: 06618/343001
; CURRENT APPLICATION NUMBER: US/09/370,098
; PRIOR FILING DATE: 1999-08-06
; PRIOR APPLICATION NUMBER: 60/095,826
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 235
; TYPE: PRT
; ORGANISM: Mus musculus
; US-09-370-098-6

Query Match 6.9%; Score 199; DB 4; Length 235;
Best Local Similarity 27.0%; Pred. No. 3.6e-12;
Matches 71; Conservative 41; Mismatches 93; Indels 58; Gaps 14;
QY 278 VSMIFLAFTIILYAFRLRSRERFKSEDAKPIHVALGSLFLNLAFLVNVGSGKSDAA 337
Db 2 ISLVCLALAIATFLLCRAVQNH-----NTYMHLCVCLFLAKILFL--TGIDKTDNQT 54
QY 338 CWARGAVFHYLLCAFTWMGLEAFHYLLA--VRVNTFYGHYFLK---LSLVGWGLPALM 393
Db 55 CAIIAGFLHYLFLACFFWMLVEAVMLFMVRNLKVVN--YFSSRNKMLHLCAFGYGLPVLV 113
QY 394 VIGTGA--NSYGLYTIIRDRENTSLELCWFREGTMYALYITVHGXYFLITFLFGM--- 447
Db 114 VIISASVQPRGYGMH-----NR-----CWLNTETG-----FIWFLGVPVCMII 151
QY 448 -VVLALVVMKIFTL-----SRATAVKRGKRNKVVLTLLGLSS--LVGVTWGLAIFT 496
Db 152 TMSVLLAWTILVLRKLCSSVSSEVSKLD-----TRLLTFKAIQAIIFGLCSWVLGIFQ 206
QY 497 --PLGLSTVYIFALFNSLQGVFI 517
Db 207 IGPLASIMAYLFTIINSLOGAFI 229

RESULT 12
US-09-370-098-4
; Sequence 4, Application US/09370098
; Patent No. 6303768
; GENERAL INFORMATION:
; APPLICANT: Benzer, Seymour
; APPLICANT: Lin, Yi-Jyun
; APPLICANT: California Institute of Technology

RESULT 14
US-08-981-189B-11
; Sequence 11, Application US/08981189B
; Patent No. 6214797
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: UROCORTIN PEPTIDES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FITCH, EVEN, TABIN & FLANNERY
; STREET: 120 S. LaSalle Street, Suite 1600
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/981,189B
; FILING DATE: 10-DEC-1997
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 60/028,144
; FILING DATE: 13-JUN-1995
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 60/002,223
; FILING DATE: 11-AUG-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Schumann, James J.
; REGISTRATION NUMBER: 20,856
; REFERENCE/DOCKET NUMBER: 57611
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 858-552-1311
; TELEFAX: 858-552-0095

INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 431 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Protein
LOCATION: 1..431
OTHER INFORMATION: /note= "Product-mouse heart derived
OTHER INFORMATION: CRF-R2 Long Form"
PUBLICATION INFORMATION:
AUTHORS: Perrin, Marilyn
AUTHORS: Donaldson, Cynthia
AUTHORS: Chen, Ruoping
AUTHORS: Blount, Amy
AUTHORS: Berggren, Travis
AUTHORS: Bileziklian, Louise
AUTHORS: Sawchenko, Paul
AUTHORS: Vale, Willie
TITLE: Identification of a second
TITLE: corticotropin-releasing factor receptor gene and
TITLE: characterization of a cDNA expressed in heart
JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
VOLUME: 92
PAGES: 2969-2973
DATE: March-1995
US-08-981-189B-11

Query Match 6.2% Score 177.5; DB 4; Length 431;
Best Local Similarity 27.8%; Pred. No. 1.4e-09;
Matches 73; Conservative 41; Mismatches 118; Indels 31; Gaps 11;
QY 271 ISQAGCGVSMIFLAFTIILYAFRLSRERFKSEDAPKIHVALGSLFLLNLA-FLVNV-- 327
DB 139 VNYLGHCVSVVALVAFLFLVLRISIRCL-----RNVIHNLITTFILRNIAFWLLQLID 193
QY 328 GSGKSGSDAACWARGAVFHYFLLCFTWGLFAFHYLLAVRVNTYFGHYFLKLSLVGW 387
DB 194 HEVHEGNEVWCRCITTFINFYVVTFFWFMFVEGCVLHTAIVMTYSTEHLRKWLFL-FIGW 252
QY 388 GLPALMVGTSANSYGLYITRDRENTSLELCWFREGTTMYALYITVHGVELITFLFGM 447
DB 253 CIPCPITIAWAVGKLY-----YEN-----EQCFWKEAGDLVDYI-YQGPVMLVLLINF 300
QY 448 VVALVVMKIFTLSRATAVKRGNKRVKLVTLGLSSLVGVTWGLAIFTP--LGLSTVYI 505
DB 301 VFELNIVRILMTKLRASTTSETIOYRKAVKATLVLLPLLGITYMLFFVNPGEDDLSQL-V 359
QY 506 FALFN---SLOGVFIC---CWF 521
DB 360 FIYFNSFLQSFQGFVSVFYCYFF 382

RESULT 15
US-08-381-433A-2
Sequence 2, Application US/08381433A
Patent No. 5786203
GENERAL INFORMATION:
APPLICANT: Lovenberg, Timothy W.
APPLICANT: Oltersdorf, Tilman
APPLICANT: Liaw, Chen
APPLICANT: Grigoriadis, Dimitri E.
APPLICANT: DeSouza, Errol B.
TITLE OF INVENTION: CORTICOTROPIN RELEASING FACTOR 2
TITLE OF INVENTION: RECEPTORS
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESS: SEED and BERRY
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington

COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/381,433A
FILING DATE: 31-JAN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mcmasters, David D.
REGISTRATION NUMBER: 33,963
REFERENCE/DOCKET NUMBER: 690068.401C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEEDANDBERRY
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 431 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-381-433A-2

Query Match 6.1% Score 174.5; DB 1; Length 431;
Best Local Similarity 27.4%; Pred. No. 2.9e-09;
Matches 72; Conservative 41; Mismatches 119; Indels 31; Gaps 11;
QY 271 ISQAGCGVSMIFLAFTIILYAFRLSRERFKSEDAPKIHVALGSLFLLNLA-FLVNV-- 327
DB 139 INYLGHCVSVVALVAFLFLVLRISIRCL-----RNVIHNLITTFILRNITWFLQLID 193
QY 328 GSGKSGSDAACWARGAVFHYFLLCFTWGLFAFHYLLAVRVNTYFGHYFLKLSLVGW 387
DB 194 HEVHEGNEVWCRCVTTIFNYFVVTFFWFMFVEGCVLHTAIVMTYSTEHLRKWLFL-FIGW 252
QY 388 GLPALMVGTSANSYGLYITRDRENTSLELCWFREGTTMYALYITVHGVELITFLFGM 447
DB 253 CIPCPITIAWAVGKLY-----YEN-----EQCFWKEGDLVDYI-YQGPILVLLINF 300
QY 448 VVALVVMKIFTLSRATAVKRGNKRVKLVTLGLSSLVGVTWGLAIFTP--LGLSTVYI 505
DB 301 VFELNIVRILMTKLRASTTSETIOYRKAVKATLVLLPLLGITYMLFFVNPGEDDLSQL-V 359
QY 506 FALFN---SLOGVFIC---CWF 521
DB 360 FIYFNSFLQSFQGFVSVFYCYFF 382

Search completed: June 11, 2002, 22:25:18
Job time: 3235 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 11, 2002, 21:27:38 ; Search time 44.13 Seconds
(without alignments)
1195.401 Million cell updates/sec

Title: US-09-733-387-44
Perfect score: 2878
Sequence: 1 MATPRGLGALLLLPTSG.....STTVSSSTARLDQAHSQEQ 549
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 segs, 96089334 residues
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	383.5	13.3	1230	2	CL3AB protein - ra
2	383.5	13.3	1231	2	latrophilin-3, spl
3	383.5	13.3	1273	2	CL3AC protein - ra
4	383.5	13.3	1274	2	latrophilin-3, spl
5	383.5	13.3	1298	2	CL3BB protein - ra
6	383.5	13.3	1299	2	latrophilin-3, spl
7	383.5	13.3	1341	2	CL3BC protein - ra
8	383.5	13.3	1342	2	latrophilin-3, spl
9	383.5	13.3	1459	2	CL3AA protein - ra
10	383.5	13.3	1503	2	latrophilin-3, spl
11	383.5	13.3	1527	2	CL3BA protein - ra
12	383.5	13.3	1571	2	latrophilin-3, spl
13	379	13.2	1240	2	latrophilin-3, spl
14	379	13.2	1283	2	latrophilin-3, spl
15	379	13.2	1308	2	latrophilin-3, spl
16	379	13.2	1351	2	latrophilin-3, spl
17	379	13.2	1512	2	latrophilin-3, spl
18	379	13.2	1550	2	latrophilin-3, spl
19	379	13.2	1580	2	latrophilin-3, spl
20	370.5	12.9	1467	2	alpha-latrophilin r
21	370.5	12.9	1472	2	latrophilin-1, bra
22	367	12.8	1420	2	latrophilin-1, bra
23	367	12.8	1452	2	CL2AB protein - ra
24	367	12.8	1463	2	CL2AA protein - ra
25	365.5	12.7	1466	2	CL2AC protein - ra
26	365.5	12.7	1471	2	CL1AA protein - ra
27	365.5	12.7	1510	2	CL1BA protein - ra
28	365.5	12.7	1515	2	CL1AB protein - ra
29	359.5	12.5	1354	2	latrophilin-2 (spl)

30	359.5	12.5	1397	2	T18377	latrophilin-2 (spl)
31	359.5	12.5	1420	2	T18385	latrophilin-2 (spl)
32	359.5	12.5	1463	2	T18386	latrophilin-2 (spl)
33	355.5	12.4	1435	2	T46611	CL2BB protein - ra
34	355.5	12.4	1467	2	T17160	CL2BA protein - ra
35	355.5	12.4	1478	2	T17185	CL2BC protein - ra
36	353	12.3	1341	2	T18301	latrophilin-2, spl
37	353	12.3	1384	2	T18366	latrophilin-2, spl
38	353	12.3	1407	2	T18381	latrophilin-2 (spl)
39	353	12.3	1450	2	T18382	latrophilin-2 (spl)
40	351	12.2	1487	2	T14324	alpha-latrophilin r
41	348	12.1	1369	2	T18379	latrophilin-2 (spl)
42	348	12.1	1412	2	T18380	latrophilin-2 (spl)
43	348	12.1	1435	2	T18387	latrophilin-2 (spl)
44	348	12.1	1478	2	T18388	latrophilin-2 (spl)
45	341.5	11.9	1356	2	T18367	latrophilin-2, spl

ALIGNMENTS

RESULT 1

T17187

CL3AB protein - rat

C:Species: Rattus norvegicus (Norway rat)

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000

C:Accession: T17187

R:Sugita, S.; Ichchenko, K.; Khvotchev, M.; Sudhof, T.C.

submitted to the EMBL Data Library, July 1998

A:Description: CL family.

A:Reference number: Z18712

A:Accession: T17187

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1230 <SUG>

A:Cross-references: EMBL:AF081155; NID:g3695136; PID:g3695137; PIDN:AAC62661.1

C:Superfamily: alpha-latrophilin receptor, calcium-independent

Query Match 13.3%; Score 383.5; DB 2; Length 1230;
Best Local Similarity 27.6%; Pred. No. 2.5e-23;
Matches 135; Conservative 77; Mismatches 182; Indels 95; Gaps 25;

QY	77	HLMKGLTKVNTPLKALVQLST-NTAEDFYFSLPS-----QVPQVQKDEKPPDR	130
DB	682	NLLKTDIVRE-NTDNIQLEVARLSTEGNLEDKFPENTGHSTIQLSANTLKQN-----	734
QY	131	VRLPKSLFSLPGRSVVRLAVTIL-DIGPGTFLFKGPRGLG-----DGSGVLNNRLVG	183
DB	735	-----GRNGEIRVAPVLYNNLGPYLSTENASKLGTAMSTNHSVIVNSPVIT	782
QY	184	LSVGOMHVTK--LAEPLEIVFSH-QRPPNMTLTCTVWDVTGKT-TGDMSEGGG-TEVR	238
DB	783	AAINTEFKNKVLADPVVTVTKHKIKQSEENFPNCSFWSKRTMTGYWSTGCRLLTTN	842
QY	239	PBGTCCCDHLTFPALLLR--PTLDQSTVH--ILTRISQAGCGVSMIFLATIILYAFLR	294
DB	843	KHTTCSNHLTNFAVLMAHVEVKHSDAHDLLDVTWGLLSLVCCLLCITFFCFF--	901
QY	295	LSRERFKSEDAPIKHAVALGSLFLNLAFVLNVGSGSGSDAACWARGAVFHYELCAFT	354
DB	902	-----RGLQSDRNTIHNKLCISLFAVELLFL--IGINRTDQPIACAVFAALLHFFFLAFT	955
QY	355	WMGLEAFHLYLLAVRVFTYFGH--YFLKLSLVGWLGPALMVIPTGTSAN--SYGLYTIRD	410
DB	956	WMFLEGVOLYIMLVEVFESHSRRKIFY--LVGYGMPALIVAVSAADYRSYG-----	1006
QY	411	RENRTSLCLCFRETTWYALYIVHGYFLITF-----LFCGMVVALVWVKIFLSR	462
DB	1007	-----TDKVCWRLDT-----YFWSFTGPTLIIMLNVIIFGLYALYKMF---H	1047
QY	463	ATAV--KERG--KNRKK-VLTLLGLSLVGVTW--GLAIFTPLGLSTVYIFALFNSLQGV	515

Db	1048	HTAILPESGCLDNIXSWVIGAIALCLGLTWAFGLMYNESTVIMAYLFTIFNSLQGM	1107
Qy	516	FICCWETIL	524
Db	1108	FIFIFHCVL	1116

RESULT 2

Tl18390
latrophilin-3, splice variant abag, brain-specific - bovine
C/Species: Bos primigenius taurus (cattle)
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C/Accession: Tl18390
R/Matsushita H.; Lelianova, V.G.; Ushkaryov, Y.A.
FEBS Lett. 443, 348-352, 1999
A/Title: The latrophilin family: multiply spliced G protein-coupled receptors
A/Reference number: Z18869; MUID:99148828
A/Accession: Tl18390
A/Status: preliminary;
A/Status: translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-1231 <MAT>
A/Cross-references: EMBL:AF111086; NID:g4164054; PID:g4164055; PIDN:AAD0532
C/Superfamily: alpha-latrotoxin receptor, calcium-independent
C/Keywords: alternative splicing; G protein-coupled receptor

```

Query Match      13.3%; Score 383.5; DB 2; Length 1231;
Best Local Similarity 27.6%; Pred. No. 2.6e-23;
Matches 135; Conservative 77; Mismatches 182; Indels 95; Gaps 25

Qy 77 HLMKEGLTKQVNTFFKALVQNLST-NTAEQFYFSLEPS-----QVPRQVMKDDEKPPDR 130
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 683 NLLKTDIVRE-WFDNIOLEVARLSTEGNEDLKPPENTGHGSTQLSANTLKQN----- 735
Qy 131 VRLPKSLFRLSPGNRSVRVLAFTIL-DIGPGTLFKGPRGLG-----DGSQVLNRLVG 183
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 736 -----GRNGEIRVAEVLNNLGPYLSTENASMKLGTTEAMSTNHSVIVNSPVIT 783
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 184 LSGCOMHVTK--LAEPLEIVFESH-ORPPPNMTLTGVFWDVTKGT-TGDWSEGCs-TEVR 238
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 784 AAINKEESNKVYLADPVFTVHKIQSENFENPNPNCFSFWSYKRTWTGYWSTQCRLLTTN 843
Qy 239 PEGTVCCDHLTFALLLR--PTLDQSTVH--ILTRISQAGCGVSMIFLAFTILYAFLR 294
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 844 KTHHTCSNHLTFNFAVLMARVEVKHSDAVHDLDDLVITWVGILLSVCLLCICTFCEFF- 902
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 295 LSRERFKSDAPKIHVALGSLFLNLAFLVNVSQSGKSDAACWARGAVHFVFLLECAFT 354
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 903 -----RGLQSDRNTIHNKLCISLFAELLFL--IGNRTDQPIACAVPAALLHFFFLAFT 956
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 355 WMGLEAFHLYLLAVRVNTFYGH--YFUKLSLVGMLPALMVIQGTGSAN--SYGLXTIRD 410
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 957 WMFLEGVOLYIMLVEFESEHSRRKFVY--LVGVMGPALIVAVSAADVRSYG----- 1007
Qy 411 RENRSLSELCWPFEGTWTALYITHVGHFLITF-----LFGWVIALVWVKIFTLR 462
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1008 -----TDKVCWLRLDT-----YFINSFGTPATLIITMLNVIFGLIYALYKMF---H 1048
Qy 463 ATAV--KERG--KNRKK-VLTLLGLSSLVGVTW--GLAIFTPLGLSTVYIFALFNSLQGV 515
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1049 HTAILKPESCLDNIKSWIGAIALCLLGLTWAFGLMYINESTVIMAYLFTFNSLQGM 1108
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 516 FICQWFTIL 524
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1109 FIFIFHCVL 1117
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 3
T17188
Cl3AC protein - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
C:Accession: T17188

```

R:Sugita, S.; Ichtchenko, K.; Khvotchev, M.; Sudhof, T.C.
submitted to the EMBL Data Library, July 1998
A:Description: CL family.
A:Reference number: Z18712
A:Accession: T17188
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1273 <SUG>
A:Cross-references: EMBL:AF081156; PID:g3695139;
C:Superfamily: alpha-latrotoxin receptor, calcium-independent
PIDN:AAC62662.1

[illegible]

RESULT 4

Tl8391

latrophilin-3, splice variant abah, brain-specific - bovine

C:Species: Bos primigenius taurus (cattle)

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000

C:Accession: Tl8391

R:Matsumita, H.; Lelianova, V.G.; Ushkaryov, Y.A.
FEBS Lett. 443, 348-352, 1999

A:Title: The latrophilin family: multiply spliced G protein-coupled receptors with different

A:Reference number: Z18869; MUID:99148828

A:Accession: Tl8391

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: mRNA

A:Residues: 1-1274 <MAT>

A:Cross-references: EMBL:AF111087; NID:q4164056; PID:q4164057; PIDN:AA005323.1

C:Superfamily: alpha-latrotoxin receptor, calcium-independent

C:Keywords: alternative splicing; G protein-coupled receptor

Query Match 13.3%; Score 383.5; DB 2; Length 1274

Best Local Similarity 27.6%; Pred. No. 2.7e-23;
Matches 135; Conservative 77; Mismatches 182; Indels 95; Gaps 25;

```
QY 77 HLMKEGLTKVNTPEFLKALVQNLST-NTAEDFYFSLPS-----QVPRQVMKDEKPPDR 130
Db 683 NLLKTDIVRE-NTDNIQLEVARLSTEGNLEDKFPENTGHGSTIQLSANTLKQ-----735
QY 131 VRLPKSLFRSLPGRNSVVRVLAUTIL-DIGPGTLFKGPRILGLG-----DGSGLNRLVWG 183
Db 736 -----GRGEIRVAVFLVNNLGPYLSTENASMKLGTEAMSTNHSVIVNSPVIT 783
QY 184 LSVGMQHVTK--LAEPLEIVFESH-QRPPNNMTLTCVFDVTKGT-TGDMSSGCS-TEVR 238
Db 784 AAINKEFSNKVYLADVPVFTVKHIKQSEENFNPCNSFWYSKRTWTGYWSTQGCRLLTTN 843
QY 239 PEGTVCCDHLTFALLLR--PTLDQSTVH--ILTRISQAGCGVSMIFLAFTIILYAFLR 294
Db 844 KTHHTCSNHLTNFAVLMAHVEVKHSDAHDLLDVTWVGLLSLVCLLICIFTCFF-902
QY 295 LSRERFKSEDPKIHVALGGSFLNLNLAFLVNVGSGSGSDAACWARGAVFHYFLCAFT 354
Db 903 -----RGLQSDRNTIHKNLICISLFAELLFL--IGINRTDQPIACAVFAALLHFFFLAAFT 956
QY 355 WMGLEAFHLYLLAVRVNTFYGH--YFLKLSLVGWLGPALMVIGTGSAN--SYGLYTIRD 410
Db 957 WMFLEGVQLYIMLVEFSEHSRRKYFY---LVGYGMPALIVAVSAADVRSYG-----1007
QY 411 RENRTSLELCWFRGTTMYALITVHGYPFLITF-----LFGMVVLALVVMKIFTLSR 462
Db 1008 -----TDKVCWLRLDT-----YFWSFIGPATLIIMLVIFGLIALYKMF---H 1048
QY 463 ATAV--KERG--KNRKK-VLTLLGSLVGVTV--GLAIFTPLGLSTVYIFALFNSLQGV 515
Db 1049 HTAILKPESGLDNIKSWIGAIALLCLGLTWAFGLMYNESTVIMAYLFTFNSLQGM 1108
QY 516 FICCWFTIL 524
Db 1109 FIFIFHCVL 1117

RESULT 5
T17199
CL3BB protein - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
C:Accession: T17199
R:Sugita, S.; Ichchenko, K.; Khvotchev, M.; Sudhof, T.C.
submitted to the EMBL Data Library, July 1998
A:Description: CL family.
A:Reference number: Z18712
A:Accession: T17199
A:Status: preliminary; translated from GB/EMBL/DBDJB
A:Molecule type: mRNA
A:Residues: 1-1298 <SUG>
A:Cross-references: EMBL:AF081158; NID:g3695142; PID:g3695143; PIDN:AACG2664.1
C:Superfamily: alpha-latotoxin receptor, calcium-independent
```

Query Match 13.3%; Score 383.5; DB 2; Length 1298;
Best Local Similarity 27.6%; Pred. No. 2.7e-23;
Matches 135; Conservative 77; Mismatches 182; Indels 95; Gaps 25;

```
QY 77 HLMKEGLTKVNTPEFLKALVQNLST-NTAEDFYFSLPS-----QVPRQVMKDEKPPDR 130
Db 750 NLLKTDIVRE-NTDNIQLEVARLSTEGNLEDKFPENTGHGSTIQLSANTLKQ-----802
QY 131 VRLPKSLFRSLPGRNSVVRVLAUTIL-DIGPGTLFKGPRILGLG-----DGSGLNRLVWG 183
Db 803 -----GRGEIRVAVFLVNNLGPYLSTENASMKLGTEAMSTNHSVIVNSPVIT 850
QY 184 LSVGMQHVTK--LAEPLEIVFESH-QRPPNNMTLTCVFDVTKGT-TGDMSSGCS-TEVR 238
Db 851 AAINKEFSNKVYLADVPVFTVKHIKQSEENFNPCNSFWYSKRTWTGYWSTQGCRLLTTN 910
```

```
QY 239 PEGTVCCDHLTFALLLR--PTLDQSTVH--ILTRISQAGCGVSMIFLAFTIILYAFLR 294
Db 911 KTHHTCSNHLTNFAVLMAHVEVKHSDAHDLLDVTWVGLLSLVCLLICIFTCFF-969
QY 295 LSRERFKSEDPKIHVALGGSFLNLNLAFLVNVGSGSGSDAACWARGAVFHYFLCAFT 354
Db 970 -----RGLQSDRNTIHKNLICISLFAELLFL--IGINRTDQPIACAVFAALLHFFFLAAFT 1023
QY 355 WMGLEAFHLYLLAVRVNTFYGH--YFLKLSLVGWLGPALMVIGTGSAN--SYGLYTIRD 410
Db 1024 WMFLEGVQLYIMLVEFSEHSRRKYFY---LVGYGMPALIVAVSAADVRSYG-----1074
QY 411 RENRTSLELCWFRGTTMYALITVHGYPFLITF-----LFGMVVLALVVMKIFTLSR 462
Db 1075 -----TDKVCWLRLDT-----YFWSFIGPATLIIMLVIFGLIALYKMF---H 1115
QY 463 ATAV--KERG--KNRKK-VLTLLGSLVGVTV--GLAIFTPLGLSTVYIFALFNSLQGV 515
Db 1116 HTAILKPESGLDNIKSWIGAIALLCLGLTWAFGLMYNESTVIMAYLFTFNSLQGM 1175
QY 516 FICCWFTIL 524
Db 1176 FIFIFHCVL 1184
```

RESULT 6

```
T18398
latrophilin-3, splice variant bbag, brain-specific - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T18398
R:Matsushita, H.; Lelianaova, V.G.; Ushkaryov, Y.A.
FEBS Lett. 443, 348-352, 1999
A:Title: The latrophilin family: multiply spliced G protein-coupled receptors with di
A:Reference number: Z18869; MUID:99148828
A:Accession: T18398
A:Status: preliminary; translated from GB/EMBL/DBDJB
A:Molecule type: mRNA
A:Residues: 1-1299 <MAT>
A:Cross-references: EMBL:AF111092; NID:g4164066; PID:g4164067; PIDN:AAD05328.1
C:Superfamily: alpha-latotoxin receptor, calcium-independent
C:Keywords: alternative splicing; G protein-coupled receptor
```

Query Match 13.3%; Score 383.5; DB 2; Length 1299;
Best Local Similarity 27.6%; Pred. No. 2.7e-23;
Matches 135; Conservative 77; Mismatches 182; Indels 95; Gaps 25;

```
QY 77 HLMKEGLTKVNTPEFLKALVQNLST-NTAEDFYFSLPS-----QVPRQVMKDEKPPDR 130
Db 751 NLLKTDIVRE-NTDNIQLEVARLSTEGNLEDKFPENTGHGSTIQLSANTLKQ-----803
QY 131 VRLPKSLFRSLPGRNSVVRVLAUTIL-DIGPGTLFKGPRILGLG-----DGSGLNRLVWG 183
Db 804 -----GRGEIRVAVFLVNNLGPYLSTENASMKLGTEAMSTNHSVIVNSPVIT 851
QY 184 LSVGMQHVTK--LAEPLEIVFESH-QRPPNNMTLTCVFDVTKGT-TGDMSSGCS-TEVR 238
Db 852 AAINKEFSNKVYLADVPVFTVKHIKQSEENFNPCNSFWYSKRTWTGYWSTQGCRLLTTN 911
QY 239 PEGTVCCDHLTFALLLR--PTLDQSTVH--ILTRISQAGCGVSMIFLAFTIILYAFLR 294
Db 912 KTHHTCSNHLTNFAVLMAHVEVKHSDAHDLLDVTWVGLLSLVCLLICIFTCFF-970
QY 295 LSRERFKSEDPKIHVALGGSFLNLNLAFLVNVGSGSGSDAACWARGAVFHYFLCAFT 354
Db 971 -----RGLQSDRNTIHKNLICISLFAELLFL--IGINRTDQPIACAVFAALLHFFFLAAFT 1024
QY 355 WMGLEAFHLYLLAVRVNTFYGH--YFLKLSLVGWLGPALMVIGTGSAN--SYGLYTIRD 410
Db 1025 WMFLEGVQLYIMLVEFSEHSRRKYFY---LVGYGMPALIVAVSAADVRSYG-----1075
```

QY 411 RENRTSLELCWFREGTMYALYITVHGYYFLITF-----LFGMVVLALVVMKIFTLR 462
Db 1076 -----TDKVCWLRDLT-----YFIWSFGPATLIIMLVIFIGIALYKMF---H 1116
QY 463 ATAV--KERG--KNRKK-VLTLLGLSSLVGVW--GLAIFTPLGLSTVYIFALFNSLQGV 515
Db 1117 HTAILKPESGCLDNKSWVIGAIALLCLLGLTWAFGLMYINESTVIMAYLFTIENSLOQM 1176
QY 516 FICCWFTIL 524
Db 1177 FIFIFHCVL 1185
RESULT 7
CL3BC protein - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T17200
R:Sugita, S.; Ichtchenko, K.; Khvotchev, M.; Sudhof, T.C.
submitted to the EMBL Data Library, July 1998
A:Description: CL family,
A:Reference number: Z18712
A:Accession: T17200
A>Status: preliminary; translated from GB/EMBL/DBDJB
A:Molecule type: mRNA
A:Residues: 1-1341 <SUG>
A:Cross-references: EMBL:AF081159; NID:g3695144; PID:g3695145; PIDN:AAC62665.1
C:Superfamily: alpha-latrotoxin receptor, calcium-independent
Query Match 13.3%; Score 383.5; DB 2; Length 1341;
Best Local Similarity 27.6%; Pred. No. 2.8e-23;
Matches 135; Conservative 77; Mismatches 182; Indels 95; Gaps 25;
QY 77 HLMKEGTLQKVTNPFKALYQNLST-NTAEDFYFSLEPS-----QVPROVMKDEKPPDR 130
Db 750 NLLKTDIVRE-NTDNIQLEVARLSTEGNEDLKFPENTGHGSTIQLSANTLQON----- 802
QY 131 VRLPKSLFRSLPGNRSVRLAVTIL-DIGPGTLFKPRIGLG-----DGSGLNLRV 183
Db 803 -----GRNGEIRVAFVLYNNLGPYLSSTENASMKLGTEAMSTNHSVIVNSPVIT 850
QY 184 LSVGMHVTK--LAEPLEIVFESH-ORPPNMTLTCVFDWTKGT-TGDRSSECCS-TEVR 238
Db 851 AAINKEFSNKVYLADPVVTVKHKOSEENFNPNCSFWSYKRTMTGYWSTOGRLLTTN 910
QY 239 PEGTVCCDHLTFALLR--PTLDQSTVH--ILTRISQAGCGVSMIFLAFTIILYAFRL 294
b 911 KTHITCSNHLTNFVLMHAEVVKHSDAHDLLDVIWVGILLSLVCLLICIFTCFF- 969
QY 295 LSREPKSEDAPIHVALGSLFLNLAFVNVGSGKSDAACWARGAVHYFLCAFT 354
Db 970 -----RGLQSDRNTHKNLCISLFAELLFL--IGINRTDQPIACAVFAALLHFFFLAAFT 1023
QY 355 WMGLEAFHLYLLAVRVNTYFEGH--YFLKLSLVGHLGALPMVIGTGSAN--SYGLYTIRD 410
Db 1024 WMFLEGVQLYIMLVFESEHSRRKIFY---LVGMPALIVAVSAADYRSYG----- 1074
QY 411 RENRTSLELCWFREGTMYALYITVHGYYFLITF-----LFGMVVLALVVMKIFTLR 462
Db 1075 -----TDKVCWLRDLT-----YFIWSFGPATLIIMLVIFIGIALYKMF---H 1115
QY 463 ATAV--KERG--KNRKK-VLTLLGLSSLVGVW--GLAIFTPLGLSTVYIFALFNSLQGV 515
Db 1116 HTAILKPESGCLDNKSWVIGAIALLCLLGLTWAFGLMYINESTVIMAYLFTIENSLOQM 1175
QY 516 FICCWFTIL 524
Db 1176 FIFIFHCVL 1184
RESULT 8

T18405
latrophilin-3, splice variant bbah, brain-specific - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T18405
R:Matsumita, H.; Lellianova, V.G.; Ushkaryov, Y.A.
FEBS Lett. 443, 348-352, 1999
A:Title: The latrophilin family: multiply spliced G protein-coupled receptors with di
A:Reference number: Z18869; MUID:99148828
A:Accession: T18405
A>Status: preliminary; translated from GB/EMBL/DBDJB
A:Molecule type: mRNA
A:Residues: 1-1342 <MAT>
A:Cross-references: EMBL:AF111093; NID:g4164068; PID:g4164069; PIDN:AAD05329.1
C:Superfamily: alpha-latrotoxin receptor, calcium-independent
Query Match 13.3%; Score 383.5; DB 2; Length 1342;
Best Local Similarity 27.6%; Pred. No. 2.8e-23;
Matches 135; Conservative 77; Mismatches 182; Indels 95; Gaps 25;
QY 77 HLMKEGTLQKVTNPFKALYQNLST-NTAEDFYFSLEPS-----QVPROVMKDEKPPDR 130
Db 751 NLLKTDIVRE-NTDNIQLEVARLSTEGNEDLKFPENTGHGSTIQLSANTLQON----- 803
QY 131 VRLPKSLFRSLPGNRSVRLAVTIL-DIGPGTLFKPRIGLG-----DGSGLNLRV 183
Db 804 -----GRNGEIRVAFVLYNNLGPYLSSTENASMKLGTEAMSTNHSVIVNSPVIT 851
QY 184 LSVGMHVTK--LAEPLEIVFESH-ORPPNMTLTCVFDWTKGT-TGDRSSECCS-TEVR 238
Db 852 AAINKEFSNKVYLADPVVTVKHKOSEENFNPNCSFWSYKRTMTGYWSTOGRLLTTN 911
QY 239 PEGTVCCDHLTFALLR--PTLDQSTVH--ILTRISQAGCGVSMIFLAFTIILYAFRL 294
Db 912 KTHITCSNHLTNFVLMHAEVVKHSDAHDLLDVIWVGILLSLVCLLICIFTCFF- 970
QY 295 LSREPKSEDAPIHVALGSLFLNLAFVNVGSGKSDAACWARGAVHYFLCAFT 354
Db 971 -----RGLQSDRNTHKNLCISLFAELLFL--IGINRTDQPIACAVFAALLHFFFLAAFT 1024
QY 355 WMGLEAFHLYLLAVRVNTYFEGH--YFLKLSLVGHLGALPMVIGTGSAN--SYGLYTIRD 410
Db 1025 WMFLEGVQLYIMLVFESEHSRRKIFY---LVGMPALIVAVSAADYRSYG----- 1075
QY 411 RENRTSLELCWFREGTMYALYITVHGYYFLITF-----LFGMVVLALVVMKIFTLR 462
Db 1076 -----TDKVCWLRDLT-----YFIWSFGPATLIIMLVIFIGIALYKMF---H 1116
QY 463 ATAV--KERG--KNRKK-VLTLLGLSSLVGVW--GLAIFTPLGLSTVYIFALFNSLQGV 515
Db 1117 HTAILKPESGCLDNKSWVIGAIALLCLLGLTWAFGLMYINESTVIMAYLFTIENSLOQM 1176
QY 516 FICCWFTIL 524
Db 1177 FIFIFHCVL 1185
RESULT 9
CL3AA protein - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T17186
R:Sugita, S.; Ichtchenko, K.; Khvotchev, M.; Sudhof, T.C.
submitted to the EMBL Data Library, July 1998
A:Description: CL family,
A:Reference number: Z18712
A:Accession: T17186
A>Status: preliminary; translated from GB/EMBL/DBDJB
A:Molecule type: mRNA
A:Residues: 1-1459 <SUG>
A:Cross-references: EMBL:AF081154; NID:g3695134; PID:g3695135; PIDN:AAC62660.1

A:Reference number: Z18869; MUID:99148828

A:Accession: T18394

A>Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: mRNA

A:Residues: 1-1283 <MAT>

A:Cross-references: EMBL:AF111095; NID:g4164062; PID:g4164063; PIDN:AAD05326.1

C:Superfamily: alpha-latrotoxin receptor, calcium-independent

C:Keywords: alternative splicing; G protein-coupled receptor

Query Match 13.2%; Score 379; DB 2; Length 1283;

Best Local Similarity 26.5%; Pred. No. 6.3e-23;

Matches 131; Conservative 77; Mismatches 189; Indels 98; Gaps 22;

QY 77 HLMKEGLTKVNTPELKVQLNST-NTAEDEFYSLEPS-----OVPRQVMKDEKPPDR 130

DB 683 NLKTDIVRE-NTDNIQLEVARLSTEGNLEDLKFPENTGHGSTIQLSANTLKQN----- 735

QY 131 VRLPKSLFRSLPGNRVVRVLAFTIL-DIGPGTLFGPRLGLG-----DGSGLVNNRLVG 183

DB 736 -----GRNGEIRVAFVLYNNLGPYLSTENASMKLGTEAMSTNHSVIVNSPVIT 783

QY 184 LSVQGMHVTK--LAEPLEIVFESH-QRPPPNMTLTCVFDVTKGT-TGDWSSEGCs-TEVR 238

DB 784 AAINKEFSNKKVYLADPWFTVKHIKQSEENFNPCSFWSYKRTMTGYWSTQGCRLLTN 843

QY 239 PEGTVCCDHLTFALLLR--PTLDQSTVH--ILTRISQAGCGVSMIFLAFTIILYAFLR 294

DB 844 KTHHTCSNHLTNFAVLMHVEVKHSDAVHLLDVLITWVGILLSLVCLLICIFTCFF- 902

QY 295 LSREPKSEDAPKIHVALGSLFLNLAFVNVGSGKSDAACWARGAVFHYFLICAPT 354

DB 903 ----RGLOSDRNTIHKNLCSLFAVELFL--IGINRTDQPIACAVFAALLHFFFLAFT 956

QY 355 WMGLEAFHLYLLAVRVNTYFGH--YFLKLSLVGWLGPALMVIGTGSAN--SYGLYTIRD 410

DB 957 WMFLEGVOLYIMLVEFESEHSRRKIFY---LVGYGMPALIVAVSAADVRSYG----- 1007

QY 411 RENRTSLELCWFREGTMYALYITVHGYYFLITF-----LFGMVVLALVVMKIF---- 458

DB 1008 -----TDKVCWLRLDT-----YFIWSFIGPATLIIMLVIFGLIALYKMFHHTA 1051

QY 459 -----TLSRATAVKRGKRRKKVLTLLGLSSLVGVTVW--GLAIFTPLGLSTVYIFALF 509

DB 1052 ILKPESGCLDNINVEDNRPPIKSWVIGATALLCLLGLTWAFGLMYINESTVIMAYLFTIF 1111

QY 510 NSLQGVFICCWFTIL 524

DB 1112 NSLQGMFIFIFHCVL 1126

RESULT 15

T18408

latrophilin-3, splice variant bbbg, brain-specific - bovine

C:Species: Bos primigenius taurus (cattle)

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000

C:Accession: T18408

R:Matsushita, H.; Lelianaova, V.G.; Ushkaryov, Y.A.

FEBS Lett. 443, 348-352, 1999

A:Title: The latrophilin family: multiply spliced G protein-coupled receptors with diffe

A:Reference number: Z18869; MUID:99148828

A:Accession: T18408

A>Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: mRNA

A:Residues: 1-1308 <MAT>

A:Cross-references: EMBL:AF111095; NID:g4164072; PID:g4164073; PIDN:AAD05331.1

C:Superfamily: alpha-latrotoxin receptor, calcium-independent

C:Keywords: alternative splicing; G protein-coupled receptor

Query Match 13.2%; Score 379; DB 2; Length 1308;

Best Local Similarity 26.5%; Pred. No. 6.5e-23;

Matches 131; Conservative 77; Mismatches 189; Indels 98; Gaps 22;

QY 77 HLMKEGLTKVNTPELKVQLNST-NTAEDEFYSLEPS-----OVPRQVMKDEKPPDR 130

DB 751 NLKTDIVRE-NTDNIQLEVARLSTEGNLEDLKFPENTGHGSTIQLSANTLKQN----- 803

QY 131 VRLPKSLFRSLPGNRVVRVLAFTIL-DIGPGTLFGPRLGLG-----DGSGLVNNRLVG 183

DB 804 -----GRNGEIRVAFVLYNNLGPYLSTENASMKLGTEAMSTNHSVIVNSPVIT 851

QY 184 LSVQGMHVTK--LAEPLEIVFESH-QRPPPNMTLTCVFDVTKGT-TGDWSSEGCs-TEVR 238

DB 852 AAINKEFSNKKVYLADPWFTVKHIKQSEENFNPCSFWSYKRTMTGYWSTQGCRLLTN 911

QY 239 PEGTVCCDHLTFALLLR--PTLDQSTVH--ILTRISQAGCGVSMIFLAFTIILYAFLR 294

DB 912 KTHHTCSNHLTNFAVLMHVEVKHSDAVHLLDVLITWVGILLSLVCLLICIFTCFF- 970

QY 295 LSREPKSEDAPKIHVALGSLFLNLAFVNVGSGKSDAACWARGAVFHYFLICAPT 354

DB 971 ----RGLOSDRNTIHKNLCSLFAVELFL--IGINRTDQPIACAVFAALLHFFFLAFT 1024

QY 355 WMGLEAFHLYLLAVRVNTYFGH--YFLKLSLVGWLGPALMVIGTGSAN--SYGLYTIRD 410

DB 1025 WMFLEGVOLYIMLVEFESEHSRRKIFY---LVGYGMPALIVAVSAADVRSYG----- 1075

QY 411 RENRTSLELCWFREGTMYALYITVHGYYFLITF-----LFGMVVLALVVMKIF---- 458

DB 1076 -----TDKVCWLRLDT-----YFIWSFIGPATLIIMLVIFGLIALYKMFHHTA 1119

QY 459 -----TLSRATAVKRGKRRKKVLTLLGLSSLVGVTVW--GLAIFTPLGLSTVYIFALF 509

DB 1120 ILKPESGCLDNINVEDNRPPIKSWVIGATALLCLLGLTWAFGLMYINESTVIMAYLFTIF 1179

QY 510 NSLQGVFICCWFTIL 524

DB 1180 NSLQGMFIFIFHCVL 1194

Search completed: June 11, 2002, 22:24:34

Job time: 3416 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 11, 2002, 22:23:31 ; Search time 25.32 seconds
(without alignments)
839.535 Million cell updates/sec

Title: US-09-733-387-44

Perfect score: 2878

Sequence: 1 MATPRGLGALLLLLLPTSG.....STTVSSSTARLDQAHSAQ 549

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	312	10.8	1584	1	BAIL_HUMAN
2	311	10.8	835	1	CD97_HUMAN
3	280.5	9.7	931	1	EMR1_MOUSE
4	277	9.6	1572	1	BAIL2_HUMAN
5	270	9.4	1522	1	BAIL3_HUMAN
6	254.5	8.8	886	1	EMR1_HUMAN
7	179.5	6.2	415	1	CRF1_SHEEP
8	177.5	6.2	431	1	CRF2_MOUSE
9	176.5	6.1	415	1	CRF1_MOUSE
10	173.5	6.0	411	1	CRF2_RAT
11	172.5	6.0	415	1	CRF1_RAT
12	168	5.8	441	1	DIHR_ACHDO
13	166	5.8	444	1	CRF1_HUMAN
14	165	5.7	415	1	CRF1_XENLA
15	161.5	5.6	413	1	CRF2_XENLA
16	161.5	5.6	420	1	CRF1_CHICK
17	161	5.6	517	1	MTH_DROYA
18	158	5.5	550	1	GLP2_RAT
19	157.5	5.5	515	1	CRF2_HUMAN
20	157	5.5	515	1	MTH_DROSI
21	156.5	5.4	458	1	MTH3_DROME
22	152.5	5.3	395	1	DIHR_MANSE
23	151	5.2	514	1	MTH_DROME
24	149.5	5.2	553	1	GLP2_HUMAN
25	149	5.2	468	1	MTH4_DROME
26	148.5	5.2	447	1	VIPR_CARAU
27	148.5	5.2	546	1	PTH2_RAT
28	148.5	5.2	585	1	PTH2_DIDMA
29	146.5	5.1	478	1	CAIR_CAVPO
30	142.5	5.0	550	1	PTH2_HUMAN
31	140.5	4.9	585	1	PTTR_PIG
32	140	4.9	676	1	MTH1_DROME
33	139	4.8	593	1	PTTR_HUMAN

34 137.5 4.8 437 1 VIPS_RAT
35 136.5 4.7 464 1 CGRR_RAT
36 135.5 4.7 489 1 GLPI_MOUSE
37 135 4.7 498 1 CALR_PIG
38 134 4.7 437 1 VIPS_MOUSE
39 134 4.7 474 1 CALR_RABIT
40 132.5 4.6 461 1 CGRR_HUMAN
41 132 4.6 466 1 GIPR_HUMAN
42 130.5 4.5 463 1 GLPI_HUMAN
43 130.5 4.5 468 1 PACR_HUMAN
44 130 4.5 591 1 PTTR_MOUSE
45 130 4.5 591 1 PTTR_RAT

ALIGNMENTS

RESULT 1
ID BAIL_HUMAN STANDARD; PRT; 1584 AA.
AC Q14514;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Brain-specific angiogenesis inhibitor 1 precursor.
GN BAIL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal brain;
RX MEDLINE=98054121; PubMed=9393972;
RA Nishimori H., Shiratsuchi T., Urano T., Kimura Y., Kiyono K.,
RA Tatsumi K., Yoshida S., Ono M., Kuwano M., Nakamura Y., Tokino T.,
RT "A novel brain-specific p53-target gene, BAIL, containing
thrombospondin type 1 repeats inhibits experimental angiogenesis.";
RL Oncogene 15:2145-2150(1997).
RN [2]
RP INTERACTION WITH BAP1.
RX MEDLINE=98321173; PubMed=9647739;
RA Shiratsuchi T., Futamura M., Oda K., Nishimori H., Nakamura Y.,
Tokino T.;
RT "Cloning and characterization of BAI-associated protein 1: a PDZ
domain-containing protein that interacts with BAIL.";
RL Biochem. Biophys. Res. Commun. 247:597-604(1998).
CC -!- FUNCTION: LIKELY TO BE A POTENT INHIBITOR OF ANGIOGENESIS IN
BRAIN AND MAY PLAY A SIGNIFICANT ROLE AS A MEDIATOR OF THE P53
SIGNAL IN SUPPRESSION OF GLOBLASTOMA. MAY FUNCTION IN CELL
ADHESION AND SIGNAL TRANSDUCTION IN THE BRAIN.
CC -!- SUBUNIT: INTERACTS WITH BAP1.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. LIKELY TO BE
CONCENTRATED AT CELL-CELL ADHESION SITES.
CC -!- TISSUE SPECIFICITY: SPECIFICALLY EXPRESSED IN BRAIN. REDUCED OR NO
EXPRESSION IS OBSERVED IN SOME GLOBLASTOMA CELL LINES AND CANCER
TISSUES.
CC -!- INDUCTION: BY P53.
CC -!- DOMAIN: THE TSP1 REPEATS INHIBIT IN VIVO ANGIOGENESIS IN RAT
CORNEA INDUCED BY BEGF.
CC -!- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.
CC -!- SIMILARITY: CONTAINS 5 TSP TYPE-1 DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 GPS DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; AB005297; BAA23647.1; -
DR

RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C: TISSUE-Peritoneal cavity;
RX MEDLINE=96132946; PubMed=8550607;
RA McKnight A.J., Macfarlane A.J., Dri P., Turley L., Willis A.C.,
RA Gordon S.;
RT "Molecular cloning of F4/80, a murine macrophage-restricted cell
RT surface glycoprotein with homology to the G-protein-linked
RT transmembrane 7 hormone receptor family.";
RL J. Biol. Chem. 271:486-489(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97312684; PubMed=9169125;
RA Lin H.H., Stubbs L.J., Mucenski M.L.;
RT "Identification and characterization of a seven transmembrane hormone
RT receptor using differential display.";
RL Genomics 41:301-308(1997).
CC -1- FUNCTION: PROBABLY INVOLVED IN CELL ADHESION WITHIN TISSUES
CC AND RECEPTOR SIGNALING.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- TISSUE SPECIFICITY: IN MACROPHAGES; BUT ABSENT FROM THOSE WHICH
CC ARE LOCALIZED WITHIN T-CELL AREAS OF LYMPH NODES AND SPLEEN.
CC LOW LEVEL OF EXPRESSION ON BLOOD MONOCYTES.
CC -1- SIMILARITY: CONTAINS 7 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.
CC -1- SIMILARITY: CONTAINS 1 GPS DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X93328; CAA63720.1; -;
DR EMBL: U66888; AAC53184.1; -;
DR HSSP: P07204; IFGD.
DR GCRdb: GCR_1309; -;
DR MGD: MGI:106912; Emr1.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_CA.
DR InterPro: IPR000832; GPCR_secretin.
DR InterPro: IPR000203; PKD_cys_rich.
DR Pfam: PF00002; 7tm_2; 1.
DR Pfam: PF00008; EGF; 6.
DR Pfam: PF01825; GPS; 1.
DR PRINTS: PR01128; EMR1HORMONER.
DR SMART: SM00179; EGF_CA; 6.
DR SMART: SM00001; EGF_like; 1.
DR SMART: SM00303; GPS; 1.
DR PROSITE: PS00010; ASX_HYDROXYL; 6.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS01187; EGF_CA; 5.
DR PROSITE: PS02221; GPS; 1.
DR PROSITE: PS00650; G_PROTEIN_RECEPTOR_F2_2; 1.
DR PROSITE: PS0261; G_PROTEIN_RECEPTOR_F2_4; 1.
KW G-protein coupled receptor; Transmembrane; Receptor; Glycoprotein;
KW EGF-like domain; Repeat; Signal.
FT SIGNAL 1 27 POTENTIAL.
FT CHAIN 28 931 CELL SURFACE GLYCOPROTEIN EMR1.
FT DOMAIN 28 644 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 645 672 POTENTIAL.
FT DOMAIN 673 679 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 680 701 POTENTIAL.
FT DOMAIN 702 711 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 712 735 POTENTIAL.
FT DOMAIN 736 754 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 755 772 POTENTIAL.
FT DOMAIN 773 792 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 793 821 POTENTIAL.
FT DOMAIN 822 839 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 840 859 POTENTIAL.

FT	DOMAIN	860	874	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	875	897	POTENTIAL.
FT	DOMAIN	898	931	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	32	80	EGF-LIKE 1.
FT	DOMAIN	81	132	EGF-LIKE 2.
FT	DOMAIN	133	172	EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	173	221	EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	222	271	EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	272	318	EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	319	367	EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	532	641	GPS.
FT	SITE	506	508	CELL ATTACHMENT SITE (POTENTIAL).
FT	DISULFID	36	48	BY SIMILARITY.
FT	DISULFID	42	57	BY SIMILARITY.
FT	DISULFID	59	79	BY SIMILARITY.
FT	DISULFID	85	98	BY SIMILARITY.
FT	DISULFID	92	107	BY SIMILARITY.
FT	DISULFID	109	131	BY SIMILARITY.
FT	DISULFID	137	149	BY SIMILARITY.
FT	DISULFID	143	158	BY SIMILARITY.
FT	DISULFID	160	171	BY SIMILARITY.
FT	DISULFID	177	189	BY SIMILARITY.
FT	DISULFID	183	198	BY SIMILARITY.
FT	DISULFID	200	220	BY SIMILARITY.
FT	DISULFID	226	239	BY SIMILARITY.
FT	DISULFID	233	248	BY SIMILARITY.
FT	DISULFID	250	270	BY SIMILARITY.
FT	DISULFID	276	286	BY SIMILARITY.
FT	DISULFID	280	295	BY SIMILARITY.
FT	DISULFID	297	317	BY SIMILARITY.
FT	DISULFID	323	336	BY SIMILARITY.
FT	DISULFID	330	345	BY SIMILARITY.
FT	DISULFID	347	366	BY SIMILARITY.
FT	CARBOHYD	148	148	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	167	167	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	229	229	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	269	269	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	283	283	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	405	405	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	417	417	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	474	474	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	498	498	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	706	706	N-LINKED (GLCNAC. .) (POTENTIAL).
SQ	SEQUENCE	931 AA;	102129 MW;	52963A667E8B76B5 CRC64;

Query Match		9.7%;	Score 280.5;	DB 1;	Length 931;
Best Local Similarity		27.6%;	Pred. No. 3.4e-13;		
Matches 100;		Conservative 57;	Mismatches 152;	Indels 53;	Gaps 19;
QY	177 LNRLVGLSGVMHVTKLAEPLIEVFSHORPPNMTL-TCVFDVDTKGTGDMSSGCS-	234			
Db	556 MNSRVGGVTGKEDFSKPIIYTLQHIQPKQKSERPICVSNW-TDVEDGRWTPSGCEI	614			
QY	235 TEVRPEGTCCDHLTFALLLRP---TLDQSTVHLITRISOAGCVSMIFLAFTILYA	291			
Db	615 VEASETHVCSNRMNLAITMASGELTWFS---LYIISHVGTVISLVCIALATATFL	670			
QY	292 FLRLSREKSEDAPIKIHVALGGSLFLNLAFLVNVGSGKSDAACWARGAVHFYFLC	351			
Db	671 LCRAVQNH-----NTYMLHLCLVCLFLAKITFL--TGIDKTDNOTACAIAGFLHYFLA	723			
QY	352 AFTWMGLEAFHLYLLA--VRVFNITYGHFLK---LSLVGWGLPALMVIQTGSA--NSVG	404			
Db	724 CFFMMLVEAYMFLVMRNKLVN-YFSSRNKMLHLCAFGYGLPVLWVIISASVQPRGV	782			
QY	405 LYTIREDRENTSLELCWFREGTTMAYLYITVHGYYLITFLFGMVVALVWV----KIFTL	460			
Db	783 MH-----NR-----CWLNTETGFIWSFL---GPVCMILITNSVLLAWLWLKQLCSV	828			
QY	461 -SRATAVYKRGKRNKKVLTLLGLSS--LVGVTWGLAIFT--PLGLSTVIFALFNSLQGV	515			
Db	829 SSEVSKLKD-----TRLLTFAKAQIFILGCSWVLGFIQIGPLASIMAYLFTIINSLOGA	883			

QY 516 FI 517
Db 884 FI 885

RESULT 4
BAI2_HUMAN STANDARD; PRT; 1572 AA.
AC O60241;
DT 16-OCT-2001 (Rel. 40, Created)
DT 01-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Brain-specific angiogenesis inhibitor 2 precursor.
GN BAI2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal brain;
RX MEDLINE=98194217; PubMed=9533023;
RA Shiratsuchi T., Nishimori H., Ichise H., Nakamura Y., Tokino T.;
RT "Cloning and characterization of BAI2 and BAI3, novel genes homologous
to brain-specific angiogenesis inhibitor 1 (BAI1).";
RL Cytochrome. Cell Genet. 79:103-108(1997).
CC -!- FUNCTION: MIGHT BE INVOLVED IN ANGIOGENESIS INHIBITION.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- TISSUE SPECIFICITY: STRONGLY EXPRESSED IN BRAIN. ALSO DETECTED IN
HEART, THYMUS, SKELETAL MUSCLE, AND DIFFERENT CELL LINES.
CC -!- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.
CC -!- SIMILARITY: CONTAINS 4 TSP TYPE-1 DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 GPS DOMAIN.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
or send an email to license@isb-sib.ch).

EMBL; AB005298; BAA25362.1; -
DR MIM; 602683; -
DR InterPro; IPR000832; GPCR_secretin.
DR InterPro; IPR000203; PKD_cys_rich.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR001879; hormo_receptor.
DR Pfam; PF00002; 7tm_2; 1.
DR Pfam; PF01825; GPS; 1.
DR Pfam; PF02793; HRM; 1.
DR Pfam; PF00090; Tsp.1; 4.
DR SMART; SM00303; GPS; 1.
DR SMART; SM00008; Hormr; 1.
DR SMART; SM00209; TSP1; 4.
DR PROSITE; PS0221; GPS; 1.
DR PROSITE; PS00649; G_PROTEIN_RECEP_F2_1; FALSE_NEG.
DR PROSITE; PS00650; G_PROTEIN_RECEP_F2_2; FALSE_NEG.
DR PROSITE; PS0227; G_PROTEIN_RECEP_F2_3; 1.
DR PROSITE; PS0261; G_PROTEIN_RECEP_F2_4; 1.
DR PROSITE; PS0092; TSP1; 4.
KW G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
FT Repeat.
FT SIGNAL.
FT CHAIN.
FT DOMAIN 21 1572
FT DOMAIN 21 924
FT TRANSMEM 925 945
FT DOMAIN 946 953
FT TRANSMEM 954 974
FT DOMAIN 975 982
FT TRANSMEM 983 1003
FT DOMAIN 1004 1024

FT TRANSMEM 1025 1045
DOMAIN 1046 1066
FT TRANSMEM 1067 1087
DOMAIN 1088 1141
FT TRANSMEM 1142 1162
DOMAIN 1163 1168
FT TRANSMEM 1169 1189
DOMAIN 1190 1572
FT DOMAIN 297 351
FT DOMAIN 352 406
FT DOMAIN 407 461
FT DOMAIN 463 517
FT DOMAIN 517 911
FT DOMAIN 911 122
FT DOMAIN 122 180
FT DOMAIN 177 180
FT DOMAIN 222 225
FT DOMAIN 1303 1306
FT DOMAIN 1352 1358
FT DOMAIN 1413 1418
FT CARBOHYD 94 94
FT CARBOHYD 179 179
FT CARBOHYD 180 180
FT CARBOHYD 344 344
FT CARBOHYD 425 425
FT CARBOHYD 548 548
FT CARBOHYD 633 633
FT CARBOHYD 855 855
SQ SEQUENCE 1572 AA; 171140 MW; A3775645B77BC285 CRC64;

Query Match 9.6%; Score 277; DB 1; Length 1572;
Best Local Similarity 27.6%; Pred. No. 1.1e-12;
Matches 132; Conservative 57; Mismatches 187; Indels 102; Gaps 24;

QY 118 RQVMDKEDK-PDPRVRLPKS-LFRSLPGRNSVVRLAVTI-LDIPGPTLFGKP-----RLG-169
Db 733 RRGKMDVRSERDLFLPKFVLSLSPGKATSGAAGSPGRGPGTVPVPGHSHORLL 792
QY 170 LGDGS-----GVNRLVGLSVGMHVTKLAEPLEIVFHSORPP-----PNWTLT-- 214
Db 793 PADPDESSYFVIGAVLYRTLGLLPPPR-PPLAVTSRVMTVTVPPTQPPAEPLITVELS 851
QY 215 -----CVQVDVTK--GTTGDWSGECST-EVRPEGTCCDHLTFALLRLPTLD 261
Db 852 YIINGTTDPHCADWDYSRADASGDWDTCQLETAQATRCQOHLSTFAVLAQPKD 911
QY 262 QSTVHIILTRIS---QAGCGVSMIFLAFTIILY-AFLRLSRERKFSADAPKIHVALGGSIF 317
Db 912 -LTLELAGSPSVPLVIGCAVSCMALLTLAIYAAFWRF-----IKSE---RSIILLNFCLS 963
QY 318 LLNLAFILVNGSGSKGSDAACWARGAVFYFLCAFTWGLEAFHLYLLAV-RVFNTPYG 376
Db 964 ILASNILILVQSRVLSKGVCTMTAAFLHFFLSFSCWLTTEAMQSYLAIVGRMTRLVR 1023
QY 377 HYFLKLSLVGGLPALMV---IGTGSANSYGLYTIIRDENRTSLELCWFR-EGTTWVALY 432
Db 1024 KRELCL--GWGLPALVWVAVSGFTTKGYG-----TSSVCWLSLEGGLLYAFV 1069
QY 433 ITVHGFLITFLFGMVVALVWVKIFTLRSATAVKRGKNRKKVLTLL-----480
Db 1070 GPAAVIVLVNMLIGIIVFNKLIMAR-DGISDKKKQKRGSRCPWASLLPCSAAGVPSp 1128
QY 481 GLSS-----LVGVTVGLAIETPLGLSTVY---IFALFNSLQGVFI 517
Db 1129 LLSSASARNAMASLWSCVVLPLDLAUTWMSAVLWMTDRRSVLFOALFAVNSAQGVFI 1186

RESULT 5
BAI3_HUMAN STANDARD; PRT; 1522 AA.
ID BAI3_HUMAN
AC O60242; O60297;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)

01-MAR-2002 (Rel. 41, Last annotation update)
Brain-specific angiogenesis inhibitor 3 precursor.
BAI3 OR KIAA0550.
Homo sapiens (Human).
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A. (LONG ISOFORM).
TISSUE=Fetal brain;
MEDLINE=98194217; PubMed=9533023;
Shiratsuchi T., Nishimori H., Ichise H., Nakamura Y., Tokino T.;
"Cloning and characterization of BAI2 and BAI3, novel genes homologous
to brain-specific angiogenesis inhibitor 1 (BAI1).";
Cytogenet. Cell Genet. 79:103-108(1997).
[2]
SEQUENCE FROM N.A. (SHORT ISOFORM).
TISSUE=Brain;
MEDLINE=98290545; PubMed=9628581;
Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,
Nomura N., Ohara O.;
"Prediction of the coding sequences of unidentified human genes. IX.
The complete sequences of 100 new cDNA clones from brain which can
code for large proteins in vitro.";
DNA Res. 5:31-39(1998).
-1- FUNCTION: MIGHT BE INVOLVED IN ANGIOGENESIS INHIBITION AND
SUPPRESSION OF GLOBLASTOMA.
-1- SUBCELLULAR LOCATION: Integral membrane protein.
-1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A
SHORT FORM; MAY BE PRODUCED BY ALTERNATIVE SPLICING.
-1- TISSUE SPECIFICITY: STRONGLY EXPRESSED IN BRAIN. ALSO DETECTED IN
HEART. REDUCED EXPRESSION IS OBSERVED IN SOME GLOBLASTOMA CELL
LINES.
-1- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.
-1- SIMILARITY: CONTAINS 1 CUB DOMAIN.
-1- SIMILARITY: CONTAINS 4 TSP TYPE-1 DOMAINS.
-1- SIMILARITY: CONTAINS 1 GPS DOMAIN.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; AB005299; BAA25363.1; -;
EMBL; AB011122; BAA25476.1; -;
MIM; 602684; -;
InterPro; IPR000859; CUB.
InterPro; IPR000832; GPCR_secretin.
InterPro; IPR000203; PKD_cys_rich.
InterPro; IPR000884; TSP1.
InterPro; IPR001879; hormn_receptor.
Pfam; PF00002; 7tm_2; 1.
Pfam; PF01825; GPS; 1.
Pfam; PF02793; HRM; 1.
Pfam; PF00090; tsp_1; 4.
SMART; SM00303; GPS; 1.
SMART; SM00008; Hormr; 1.
SMART; SM00209; TSP1; 4.
PROSITE; PS01180; CUB; 1.
PROSITE; PS02221; GPS; 1.
PROSITE; PS00649; G_PROTEIN_RECEP_F2_1; FALSE_NEG.
PROSITE; PS00650; G_PROTEIN_RECEP_F2_2; FALSE_NEG.
PROSITE; PS05027; G_PROTEIN_RECEP_F2_3; 1.
PROSITE; PS02611; G_PROTEIN_RECEP_F2_4; 1.
PROSITE; PS50092; TSP1; 4.
G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
Repeat; Alternative splicing.
KW Repeat; 1 24 POTENTIAL.
FT SIGNAL 25 1522 BRAIN-SPECIFIC ANGIOGENESIS INHIBITOR 3.
FT CHAIN 25 880 EXTRACELLULAR (POTENTIAL).
FT DOMAIN

FT	TRANSMEM	881	901	1 (POTENTIAL).
FT	DOMAIN	902	910	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	911	931	2 (POTENTIAL).
FT	DOMAIN	932	939	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	940	960	3 (POTENTIAL).
FT	DOMAIN	961	981	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	982	1002	4 (POTENTIAL).
FT	DOMAIN	1003	1023	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	1024	1044	5 (POTENTIAL).
FT	DOMAIN	1045	1098	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	1099	1119	6 (POTENTIAL).
FT	DOMAIN	1120	1125	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	1126	1146	7 (POTENTIAL).
FT	DOMAIN	1147	1522	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	30	159	CUB.
FT	DOMAIN	291	344	TSP TYPE-1 1.
FT	DOMAIN	345	399	TSP TYPE-1 2.
FT	DOMAIN	400	454	TSP TYPE-1 3.
FT	DOMAIN	455	509	TSP TYPE-1 4.
FT	DOMAIN	510	868	GPS.
FT	DOMAIN	942	945	POLY-THR.
FT	DOMAIN	1173	1176	POLY-SER.
FT	CARBOHYD	51	51	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	54	54	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	82	82	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	105	105	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	241	241	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	337	337	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	418	418	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	540	540	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	625	625	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	779	779	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	812	812	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	828	828	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	937	937	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	643	665	MISSING (IN SHORT ISOFORM).
FT	VARSPPLIC	990	1007	LPALVAVTSVGTTRTKGY -> KHIDIPFHALKMLTNTH (IN SHORT ISOFORM).
FT	VARSPPLIC	1008	1522	MISSING (IN SHORT ISOFORM).
FT	SEQUENCE	1522 AA;	171490 MW;	D22D0A5D4B862502 CRC64;

Query Match 9.4%; Score 270; DB 1; Length 1522;
Best Local Similarity 25.1%; Pred. No. 3.6e-12;
Matches 117; Conservative 67; Mismatches 159; Indels 124; Gaps 23;

QY	129	DRVRLPKSLFRSLPGRSVVRVLA	VTILDITDITGPTLFGKPRGLGD	--GSGVLNRLVGLSV 186
DB	737	DRVVLPKSIPTV-SSKELDESSVFLG	--AVLYKNLDLILPTLRNYTVINSKII	--- 788
QY	187	GQHVTKLAEP	-----LEIVFESHQRPNNMTLT	--CVFWD--VTGTTGDMSSGECSTE 236
DB	789	---VVTIRPEPKTDFLEIEIAHL	---ANGTLNVCVLDWDDSKNESLGTWSTGCKTV 842	
QY	237	VREGE-TVCCDHLTFALLRP	---TLQSTVHLITRISQAGCVSMIFLAFTILYA 291	
DB	843	LTDASHTKCLCDRLSTFAILAQQP	PREIIMESGTPSVTLI--VSGLSCLALITLAVVYA 900	
QY	292	-----FLRLSRPKSEDAPKIHVALGGS	FLNLMA-----FLNVVSGSGSDAACVA 340	
DB	901	ALWRYIRSR	-----SILINFCLSISSNIIILVGTQTHNSICT 943	
QY	341	RGAVFHYELCAFTWGLEAFHLYLLAV	RVFNTRYGHVFKLSLVGWLGLPALMV---IG 396	
DB	944	TTAFLHFFFLASFVFWLTTEAWQSYMA	VTGKTRLLRKFLCL--GWSGLPALVATSVG 1000	
QY	397	TGSANSYGLYTIIRDRENRNRTSLCWF	R-EGTTMYALYITVHGTYLITFTFGMVVLAVVM 455	
DB	1001	FTRTKGYG	-----TDHYCWLISLEGLLYAFVGPAAAVLVN	VMVIGILFVFNKL- 1048
QY	456	KITLSRAPVAKRGKNRK	-----KVLTLGLSS-----	484
DB	1049	-----SRDGLDKLKHRAQOMSEPHSGL	TLCKACGCVVSTTALSATTASNANASLWSSC 1103	

Qy 177 -----LNRLVGLSVGMHVTYK----LAEPLEIVESHORPPNMTL---TCVFWDV-T 221
Db 503 TTSEIKLKMNSRVVG---GIMTEKKGDFSDP---IIVTLENVQPKQKFPICVSWSTDV 557
Qy 222 KGTGDMSSSEC-STEVREPTVCCDHLFFALLRP---TLDOSIVHILTRISQAGC 277
Db 558 KG--GRWTSFCVILEASEYTTICSCNOMANLAVIMASGELTMDFS-----LYTIISHVGII 611
Qy 278 VSMFLAFTIILYAFRLSRERKSEDAKPIHVALGGSFLFLNLAFLVNVGSGSKGSDAA 337
Db 612 ISLVCLVLAITATFLCSIRNH-----NTYLHLHLVCVLLAKTLFL--AGIHKTDNKTG 664
Qy 338 CWARGAVHFLCAFTWMLGLEAPHYLLA--VRVFNVTYGHYFLKLSLV---GWGLPAL 392
Db 665 CAIITAGLHLYFLACFFMVLVEAVILFVLRNKLKVN--YESSRNKMLHICAFYGLPML 723
Qy 393 MVICTGSA--NSYGLYTIRORENTSLELCWFREGITMYALYITVHGIFLITFLGMVVL 450
Db 724 VVVISASVQVQGMH-----NR-----CWLNETG-----FIWSEL-GEVCT 760
Qy 451 ALVV-----WKIFTL-SRATAVAKERGNK--KVLTLGLSL--VGVTWGLAIFT--P 497
Db 761 VIVINSLLTWTWLILQRLSSVNAEYSTLKDTRLLTFKFAQLFILGCSWVLGIFQIGP 820
Qy 498 LGLSTVIFALNSLQGVFI-----CC-----WFTILYLPQSTTVSSSTAR--L 540
Db 821 VAGVMAYLFTIINSLOGAFILHCLNGQVREYKRWITGKTKPSS-----QSTSKILL 876
Qy 541 DQAHASQ 548
Db 877 SSMPASK 884

RESULT 7
CRFL_SHEEP
ID CRFL_SHEEP STANDARD; PRT; 415 AA.
AC 062772;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Corticotropin releasing factor receptor 1 precursor (CRF-R) (CRFL)
DE (Corticotropin releasing hormone receptor 1) (CRH-R 1).
GN CRH1.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pituitary;
RX MEDLINE=99079328; PubMed=9863624;
RA Myers D.A., Trinh J.V., Myers T.R.;
RT "Structure and function of the ovine type 1 corticotropin releasing
RT factor receptor (CRFL) and a carboxyl-terminal variant.";
RL Mol. Cell. Endocrinol. 144:21-35(1998).
CC -1- FUNCTION: THIS IS A RECEPTOR FOR CORTICOTROPIN RELEASING FACTOR.
CC SHOWS HIGH-AFFINITY CRF BINDING. THE ACTIVITY OF THIS RECEPTOR IS
CC MEDIATED BY G PROTEINS WHICH ACTIVATE ADENYL CYCLASE.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- PTM: CARBOXYL-TERMINAL SER OR THR RESIDUES MAY BE PHOSPHORYLATED.
CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF054582; AAC08027.1; .

DR InterPro: IPR000832; GPCR_secretin.
DR InterPro: IPR001879; hormn_receptor.
DR Pfam: PF00002; 7tm_2; 1.
DR PRINTS: PF02793; HRM; 1.
DR PRINTS: PF00249; GPCRSECRETIN.
DR SMART: SM00008; Hormn; 1.
DR PROSITE: PS00649; G-PROTEIN_RECEP_F2_1; FALSE_NEG.
DR PROSITE: PS00650; G-PROTEIN_RECEP_F2_2; 1.
DR PROSITE: PS00651; G-PROTEIN_RECEP_F2_3; 1.
DR PROSITE: PS00652; G-PROTEIN_RECEP_F2_4; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
KW Phosphorylation.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 415 CORTICOTROPIN RELEASING FACTOR RECEPTOR
FT 1.
FT DOMAIN 25 121 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 122 142 1 (POTENTIAL).
FT DOMAIN 143 151 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 152 171 2 (POTENTIAL).
FT DOMAIN 172 189 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 190 213 3 (POTENTIAL).
FT DOMAIN 214 227 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 228 249 4 (POTENTIAL).
FT DOMAIN 250 268 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 269 291 5 (POTENTIAL).
FT DOMAIN 292 314 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 315 334 6 (POTENTIAL).
FT DOMAIN 335 349 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 350 369 7 (POTENTIAL).
FT DOMAIN 370 415 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 45 45 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 51 51 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 78 78 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 90 90 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 98 98 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 415 AA; 47558 MW; FASF652DI2B4CDC4 CRC64;

Query Match 6.2%; Score 179.5; DB 1; Length 415;
Best Local Similarity 23.5%; Pred. No. 3.9e+06;
Matches 101; Conservative 54; Mismatches 163; Indels 111; Gaps 20;
Qy 143 GNRVVRVLAFTLDIGPGLFKG-----PRGLGDG--SGVLNRLVGLS-----VG 187
Db 2 GRPQLRALKALLGLLNSISASLQDQHCESLSASVSLQCNASVDLNGTCWPQSPAG 61
Qy 188 QMHVTKLAEPLEIVESHORPPNMTLTCV--FWDVTGKTT-----GDWSEGCST 235
Db 62 QLVV-----RP-----CLVFFGVRYNTTSNGYRVCLANGTWAAR----- 96
Qy 236 EVRPEGTVCDDHLTFPALLRPTLDQSTVHILTRISQAGCVSMIFLAFTIILYAFRL 295
Db 97 -----VNHSEPCQELSEGEKSAHYHIAVILYGLHCISIALLLVAFVL--FLRL 144
Qy 296 SRRPKSEDAKPIHVALGGSFLFLNLAFLV-----NVGSGSKGSDAACWAR-- 341
Db 145 RSIRCVRN---IHNWLLISAFILRNATWVVLQMTSPVHQSNVG-----WCLRV 191
Qy 342 GAVFHYLLCAFTWMLGLEAPHYLLAVRVFNVTYGHYFLKLSLVGWLGPALMVGTSAN 401
Db 192 TAAYNFWHTVTFWFGGECYLHTAVLTYSTDRLRKWMFI--CIGWGVFFPIIYAWAIGK 250
Qy 402 SYGLTYTIRDRENTSLELCWFREGITMYALYITVHGIFLITFLGMVVLVWVKIFTL 461
Db 251 LY-----YDN-----EKCWFGKRGVYTDVI--VQGPMLVLLINFLFVILMTKL 298
Qy 462 RATAVKEGKRNKVVTLGLSSLVGVTWGLAIFTPLGLSTV--YIFALFN-----SLQGV 515
Db 299 RASTTSETIQYRKAVKATLVLLPLIGITYMLFFVNP--GEDEVSRVVFYFNFSLESFOG 357
Qy 516 FICCWFTIL 524
|: |: |

Db 358 FVSFYCFCL 366

RESULT 8
CRF2_MOUSE

ID CRF2_MOUSE STANDARD; PRT; 431 AA.
AC Q60748; Q60808; Q60783;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Corticotropin releasing factor receptor 2 precursor (CRF-R 2) (CRF2)
DE (Corticotropin-releasing hormone receptor 2) (CRH-R 2) (CRF-RB)
DE (CRH-R2).
GN CRR2 OR CRF2R.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=95224061; PubMed=7708757;
RA Perrin M., Donaldson C., Chen R., Blount A., Berggren T.,
RA Bilezikjian L., Sawchenko P., Vale W.;
RT "Identification of a second corticotropin-releasing factor receptor
RL gene and characterization of a cDNA expressed in heart.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:2969-2973(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C; TISSUE=Heart;
RX MEDLINE=95166778; PubMed=7755719;
RA Kishimoto T., Pearce R.V. II, Lin C.R., Rosenfeld M.G.;
RT "A sauvagine/corticotropin-releasing factor receptor expressed in
RT heart and skeletal muscle.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:1108-1112(1995).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C; TISSUE=Heart;
RX MEDLINE=96015396; PubMed=7565810;
RA Stenzel P., Kesterson R., Yeung W., Cone R.D., Rittenberg M.B.,
RA Stenzel-Poore M.P.;
RT "Identification of a novel murine receptor for
RT corticotropin-releasing hormone expressed in the heart.";
RL Mol. Endocrinol. 9:637-645(1995).
CC -!- FUNCTION: THIS IS A RECEPTOR FOR CORTICOTROPIN RELEASING FACTOR.
CC SHOWS HIGH-AFFINITY CRF BINDING. THE ACTIVITY OF THIS RECEPTOR IS
CC MEDIATED BY G PROTEINS WHICH ACTIVATE ADENYLYL CYCLASE.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN THE HEART. ALSO EXPRESSED
CC IN LUNGS, SKELETAL MUSCLE, GASTROINTESTINAL TRACT, EPIDIDYMS, AND
CC BRAIN.
CC -!- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; U17858; AAA68026.1; -;
CC DR EMBL; U21729; AAC52174.1; -;
CC DR EMBL; U19939; AAC52243.1; -;
CC DR GCRDb; GCR_1705; -;
CC DR GCRDb; GCR_1711; -;
CC DR MGB; MGI:894312; Cchr2.
CC InterPro: IPR000832; GPCR_secretin.
CC DR InterPro: IPR001879; hormo_receptor.
CC DR Pfam; PF00002; 7tm_2; 1.
CC DR Pfam; PF02793; HRM; 1.
CC DR PRINTS; PF00249; GPCRSECRETIN.
CC SMART; SM00008; HORMR; 1.

DR PROSITE; PS00649; G_PROTEIN_RECEP_F2_1; 1.
DR PROSITE; PS00650; G_PROTEIN_RECEP_F2_2; 1.
DR PROSITE; PS0227; G_PROTEIN_RECEP_F2_3; 1.
DR PROSITE; PS0261; G_PROTEIN_RECEP_F2_4; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 24
FT CHAIN 25 431
FT FT
FT CORTICOTROPIN RELEASING FACTOR RECEPTOR
FT FT
FT
FT DOMAIN 25 138
FT TRANSMEM 139 159
FT DOMAIN 160 168
FT TRANSMEM 169 188
FT DOMAIN 189 205
FT TRANSMEM 206 229
FT DOMAIN 230 243
FT TRANSMEM 244 265
FT DOMAIN 266 284
FT TRANSMEM 285 307
FT DOMAIN 308 330
FT TRANSMEM 331 350
FT DOMAIN 351 365
FT TRANSMEM 366 385
FT DOMAIN 386 431
FT CARBOHYD 52 52
FT CARBOHYD 61 61
FT CARBOHYD 94 94
FT CARBOHYD 106 106
FT CARBOHYD 114 114
FT CONFLICT 3 5
FT CONFLICT 126 126
FT CONFLICT 392 393
FT CONFLICT 396 397
FT CONFLICT 408 408
SQ SEQUENCE 431 AA; 49923 MW; A6D9EDE575DB8061 CRC64;

Query Match 6.2%; Score 177.5; DB 1; Length 431;
Best Local Similarity 27.8%; Pred. No. 5.8e-06;
Matches 73; Conservative 41; Mismatches 118; Indels 31; Gaps 11;

QY 271 ISQAGCVSMIFLAFTIILYAFRLRLSRERKPSDAPKIHVALGSLFLNLAL-FLVNV-- 327
DB 139 VNYLGHCVSVVALVAFLFLVLRSLRCL-----RNVIHNLITTFILRNIAFLQLID 193
QY 328 GSGSKGSDAACWARGAVFHYFLLCAFTWGLAEFHLVLLAVRVNVEYFGHYFLKLSLVGW 387
DB 194 HEVHEGNEVWCRCITTFINFFVNTFFMFEVGCYLHTAIVMTYSTEHLRKLFL-FIGW 252
QY 388 GLPALMVIGTGSANSYGLYTIIRDRENTSLCLCFREGTMYALXITVHGYYFLITFLFGM 447
DB 253 CIPCPIIAWAGVKLY-----YEN---EQCFWFGKAGDLYDYI-YQGPVMLVLLINF 300
QY 448 VVIALVVKITFTLSRAVAVKRGKNNKKVLTLLGLSSLVGVWGLAIFTP--IGLSTVYI 505
DB 301 VFLFNIVRIILMTKLRASTSETTIQYKAVKATVLLVLLGLITMYLFFVNPGEDDLQI-V 359
QY 506 FALFN----SLOGVFVIC---CWF 521
DB 360 FIYFNFLQSFQGFVSVFYCF 382

RESULT 9
CRFL_MOUSE
ID CRFL_MOUSE STANDARD; PRT; 415 AA.
AC P35347;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Corticotropin releasing factor receptor 1 precursor (CRF-R) (CRFL)
DE (Corticotropin-releasing hormone receptor 1) (CRH-R 1).
GN CRHRI OR CRHR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

GCRdb; GCR_1415; .
 GCRdb; GCR_1469; .
 DR InterPro; IPR000832; GPCR_secretin.
 DR Pfam; PF00002; 7tm_2; 1.
 DR PRINTS; PR00249; GPCRSECRETIN.
 DR SMART; SM00008; Hormr; 1.
 DR PROSITE; PS00649; G_PROTEIN_RECP_F2_1; 1.
 DR PROSITE; PS00650; G_PROTEIN_RECP_F2_2; 1.
 DR PROSITE; PS00227; G_PROTEIN_RECP_F2_3; 1.
 DR PROSITE; PS00261; G_PROTEIN_RECP_F2_4; 1.
 KW G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
 KW phosphorylation.
 FT SIGNAL 1 24 POTENTIAL.
 FT CHAIN 25 415 CORTICOTROPIN RELEASING FACTOR RECEPTOR
 FT 1.
 FT DOMAIN 25 121 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 122 142 1 (POTENTIAL).
 FT DOMAIN 143 151 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 152 171 2 (POTENTIAL).
 FT DOMAIN 172 189 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 190 213 3 (POTENTIAL).
 FT DOMAIN 214 227 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 228 249 4 (POTENTIAL).
 FT DOMAIN 250 268 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 269 291 5 (POTENTIAL).
 FT DOMAIN 292 314 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 315 334 6 (POTENTIAL).
 FT DOMAIN 335 349 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 350 369 7 (POTENTIAL).
 FT DOMAIN' 370 415 CYTOPLASMIC (POTENTIAL).
 FT CARBOHYD 32 32 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 38 38 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 45 45 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 78 78 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 90 90 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 98 98 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT SEQUENCE 415 AA; 47842 MW; 48D6704B31D4C013 CRC64;
 Query Match 6.0%; Score 172.5; DB 1; Length 415;
 Best Local Similarity 24.9%; Pred. No. 1.3e-05;
 Matches 102; Conservative 57; Mismatches 178; Indels 73; Gaps 21;
 QY 143 GNRSVRL--AVTILDIGP--GTLFQGR---LGL-GDGSVLLNRLVGLSVGMHVTKLA 195
 DB 2 GRPQLRLVALLLLGNPVSTSLQDRCENLSLTVSGLOCNASVDL-IGTCWPRSPA 60
 QY 196 EPLVIFSHORPPPNMTLTCTVFDVDTKGT-----GDWSSEGGCGSTEVRPEGTCC 245
 DB 61 GOLVV-----RCPD-----AFFGVRYNTNNGYRECLANGSWAARVNYSECQ----- 103
 QY 246 CDHITFFALLRLPTDOSTVHITRISQACGVSMFLAFTILYFLRLSRERPKSEDA 305
 DB 104 -----EILNEEKSKVHVHVAIYNYLGHICISLVALLVAFVL--FLRLSIRCLRN-- 152
 QY 306 PKHVALGGSLFLINLAFLV---NVGSGSKGSDAACWAR--GAVFHYFLLCFTWNGLEA 360
 DB 153 -IIHWNLSIFILRNATFWVQVLTSPVHQSNVA-WRLVTAAYNFHVTFWFMFGEG 210
 QY 361 FHYLLAVRVNRYFHYFKLSLVGWLPLAMVIGTGSANSYGLYITDRNRTSLKLC 420
 DB 211 CYLHTATVLVYSDRLRKWN-FVCIQGWGPFPIIIVAWAIGKLH-----YDN-----EKC 258
 QY 421 WFRGTTMYALYITVHGYSFLITFLFGMVVALVWVKIIFLSRATVKGKRNKVVLL 480
 DB 259 WFGRRPGVYTDYI-YQGPMLVLLINFIPLFNIVRLMTKLRASTSETIQYRKAVKATL 317
 QY 481 GLSLVGVTVWGLAIFPLGLSTV--YIPALFN-----SLOGVFIQCVFTIL 524
 DB 318 VLLPFLGITYMLFVNP-GEDEVSRVVFYIFNSFLESFQGFVSVFYCF 366
 RESULT 12
 DIHR_ACHDO STANDARD; PRT; 441 AA.
 AC Q16983;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Diuretic hormone receptor precursor (DH-R).
 OS Acheta domestica (House cricket).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Orthoptera; Orthopteroidea; Gryllidae;
 OC Gryllinae; Acheta.
 OX NCBI_TaxID=6997;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Malpighian tubules;
 RX MEDLINE=96189577; PubMed=8673074;
 RA Reagan J.D.;
 RT "Molecular cloning and function expression of a diuretic hormone
 receptor from the house cricket, Acheta domestica.";
 RL Insect Biochem. Mol. Biol. 26:1-6(1996).
 CC -1- FUNCTION: RECEPTOR FOR THE INSECT DIURECTIC HORMONE. THE ACTIVITY
 OF THIS RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE ADENYLYL
 CYCLASE.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN MALPIGHIAN TUBULES.
 CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation ~
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See http://www.isb-sib.ch/annouce/
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U15959; AAC47000.1; .
 DR GCRdb; GCR_1380; .
 DR InterPro; IPR000832; GPCR_secretin.
 DR Pfam; PF00002; 7tm_2; 1.
 DR PRINTS; PR00249; GPCRSECRETIN.
 DR SMART; SM00008; Hormr; 1.
 DR PROSITE; PS00649; G_PROTEIN_RECP_F2_1; 1.
 DR PROSITE; PS00650; G_PROTEIN_RECP_F2_2; 1.
 DR PROSITE; PS00227; G_PROTEIN_RECP_F2_3; 1.
 DR PROSITE; PS00261; G_PROTEIN_RECP_F2_4; 1.
 KW G-protein coupled receptor; Transmembrane; Glycoprotein; Signal.
 FT SIGNAL 1 24 POTENTIAL.
 FT CHAIN 25 441 DIURETIC HORMONE RECEPTOR.
 FT DOMAIN 25 134 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 135 158 1 (POTENTIAL).
 FT DOMAIN 159 166 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 167 187 2 (POTENTIAL).
 FT DOMAIN 188 194 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 195 224 3 (POTENTIAL).
 FT DOMAIN 225 238 4 (POTENTIAL).
 FT TRANSMEM 239 260 5 (POTENTIAL).
 FT DOMAIN 261 291 6 (POTENTIAL).
 FT TRANSMEM 292 315 7 (POTENTIAL).
 FT DOMAIN 316 338 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 339 357 8 (POTENTIAL).
 FT DOMAIN 358 371 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 372 391 9 (POTENTIAL).
 FT DOMAIN 392 441 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 99 99 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 107 107 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 112 112 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT SEQUENCE 441 AA; 49538 MW; 501915AC2E776C5C CRC64;

Query Match 5.8%; Score 168; DB 1; Length 441;
 Best Local Similarity 24.6%; Pred. No. 3e-05;
 Matches 69; Conservative 50; Mismatches 115; Indels 46; Gaps 13;

QY 275 GCGVSMFIAFTIILYAF---LRLSRFRKSEDAPIHVALGSLFLNLAFLVN----- 326
 | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
 DB 140 GFCLSLVAIAVAIWIIFLYFKDLRLRN-----TIHNTLMATVICNDATWISAVQOE 191

QY 327 -VSGSGKSDAACWARGAVFHYELLCATWAGLEAFHLYLLAVRVNTFYGHYFKLS-- 383
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
 DB 192 YVNG-----GLCSVLAVLMHIFITNFMEVFEGLYFLFLVVA---TFTGEK-VKLGQIY 242

QY 384 -LVGGLPALMVIIGTSANSYGLYTIIRDNRKTS-----LELC-WFRGGTTMYALYIVH 436
 :
 DB 243 IILGWGIPGVIVTVAIIKHLG-KTAPDNAGESHPVLLIKHCPMAED-----YFDWIH 296

QY 437 GYELITFL-FGMVIALVNVKIFTLSRATAVKRGKRNKKVTLTGLGSLVGVTWGLAIF 495
 : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
 DB 297 QAPVITVLAVNLVFLFSIMWVLITKQSANTAETQYRKATKALLVLPPLGITYILMMQ 356

QY 496 TPLGLSTVYIF---ALFNSLQGVFICWFRTLYLPSTOST 531
 | : : : | : | : | : : | : : | : : | : : | : : | : : | : :
 DB 357 GPMGDVAGHVFNQAALLLSLQGFVALFYCFNLTEVQNT 396

RESULT 13
 CRF1_HUMAN STANDARD; PRT: 444 AA.
 AC P34598; Q13008; Q9UK64;
 DT 01-FEB-1994 (rel. 28, Created)
 DT 01-FEB-1994 (rel. 28, Last sequence update)
 DT 16-OCT-2001 (rel. 40, Last annotation update)
 DE Corticotropin-releasing factor receptor 1 precursor (CRF-R) (CRF1)
 DE (Corticotropin-releasing hormone receptor 1) (CRH-R 1).
 GN CRHR1 OR CRHR OR CRFR.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Pituitary; PubMed=7692441;
 RX MEDLINE=94022296; PubMed=7692441;
 RA Vita N., Laurent P., Lefort S., Chalon P., Lelias J.-M., Kaghad M.,
 le Fur G., Caput D., Ferrara P.;
 RT "Primary structure and functional expression of mouse pituitary and
 human brain corticotropin releasing factor receptors";
 RL FEBS Lett. 335:1-5(1993).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=94063063; PubMed=8243652;
 RA Vita N., Laurent P., Lefort S., Chalon P., Lelias J.-M., Kaghad M.,
 le Fur G., Caput D., Ferrara P.;
 RT "Primary structure and functional expression of mouse pituitary and
 human brain corticotropin releasing factor receptors";
 RL FEBS Lett. 335:1-5(1993).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Hippocampus;
 RX MEDLINE=95110332; PubMed=7811272;
 RA Ross P.C., Kostas C.M., Ramabhadran T.V.;
 RT "A variant of the human corticotropin-releasing factor (CRF)
 receptor: cloning, expression and pharmacology";
 RL Biochem. Biophys. Res. Commun. 205:1836-1842(1994).
 RN [5]
 RP SEQUENCE FROM N.A. (ISOFORM CRF-R4).
 RX MEDLINE=20065650; PubMed=10598591;
 RA Grammatopoulos D.K., Dai Y., Randeava H.S., Levine M.A., Karteris E.,
 Easton A.J., Hillhouse E.W.;

"A novel spliced variant of the type 1 corticotropin-releasing hormone receptor with a deletion in the seventh transmembrane domain present in the human pregnant term myometrium and fetal membranes."; Mol. Endocrinol. 13:2189-2202(1999).
 CC -1- FUNCTION: THIS IS A RECEPTOR FOR CORTICOTROPIN-RELEASING FACTOR.
 CC SHOWS HIGH-AFFINITY CRF BINDING. THE ACTIVITY OF THIS RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE ADENYLATE CYCLASE.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- ALTERNATIVE PRODUCTS: 4 ISOFORMS; CRF-R1 (SHOWN HERE), CRF-R2, CRF-R3 AND CRF-R4/1D; ARE PRODUCED BY ALTERNATIVE SPLICING. CRF-R1 APPEARS TO BE THE PREDOMINANT FORM. CRF-R3 DOES NOT BIND TO CRF WITH A HIGH AFFINITY.
 CC -1- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN THE CEREBELLUM, PITUITARY, CEREBRAL CORTEX AND OLFACTORY LOBE.
 CC -1- PTM: CARBOXYL-TERMINAL SER OR THR RESIDUES MAY BE PHOSPHORYLATED.
 CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L23333; AAA35719.1; -
 DR EMBL: L23332; AAA35718.1; -
 DR EMBL: X72304; CAA51052.1; -
 DR EMBL: AF039523; AAC69993.1; -
 DR EMBL: AF039510; AAC69993.1; JOINED.
 DR EMBL: AF039511; AAC69993.1; JOINED.
 DR EMBL: AF039512; AAC69993.1; JOINED.
 DR EMBL: AF039513; AAC69993.1; JOINED.
 DR EMBL: AF039514; AAC69993.1; JOINED.
 DR EMBL: AF039515; AAC69993.1; JOINED.
 DR EMBL: AF039516; AAC69993.1; JOINED.
 DR EMBL: AF039517; AAC69993.1; JOINED.
 DR EMBL: AF039518; AAC69993.1; JOINED.
 DR EMBL: AF039519; AAC69993.1; JOINED.
 DR EMBL: AF039520; AAC69993.1; JOINED.
 DR EMBL: AF039521; AAC69993.1; JOINED.
 DR EMBL: AF039522; AAC69993.1; JOINED.
 DR EMBL: U16273; AAC50073.1; -
 DR EMBL: AF180301; AAD52688.1; -
 DR GCRDB: GCR_0679; -
 DR GCRDB: GCR_0680; -
 DR GCRDB: GCR_0846; -
 DR GCRDB: GCR_2030; -
 DR MIM: 122561; -
 DR InterPro: IPR000832; GPCR_secretin.
 DR InterPro: IPR001879; hormn_receptor.
 DR Pfam: PF00002; 7tm_2; 1.
 DR Pfam: PF02793; HRM; 1.
 DR PRINTS; PR00249; GPCRSECRETIN.
 DR SMART; SM00008; Hormr; 1.
 DR PROSITE; PS00649; G_PROTEIN_RECP_F2_1; 1.
 DR PROSITE; PS00650; G_PROTEIN_RECP_F2_2; 1.
 DR PROSITE; PS00650; G_PROTEIN_RECP_F2_3; 1.
 DR PROSITE; PS00650; G_PROTEIN_RECP_F2_4; 1.
 DR PROSITE; PS00650; G_PROTEIN_RECP_F2_5; 1.
 DR G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
 KW Phosphorylation; Alternative splicing.
 FT SIGNAL 1 24
 FT CHAIN 25 444
 FT CORTICOTROPIN-RELEASING FACTOR RECEPTOR
 FT
 FT DOMAIN 25 121
 FT TRANSMEM 122 142
 FT DOMAIN 143 180
 FT TRANSMEM 181 200
 FT DOMAIN 201 218
 FT TRANSMEM 219 242
 FT DOMAIN 243 256
 FT TRANSMEM 257 278
 FT DOMAIN 279 297
 FT EXTRACELLULAR (POTENTIAL).
 FT 1 (POTENTIAL).
 FT CYTOPLASMIC (POTENTIAL).
 FT 2 (POTENTIAL).
 FT EXTRACELLULAR (POTENTIAL).
 FT 3 (POTENTIAL).
 FT CYTOPLASMIC (POTENTIAL).
 FT 4 (POTENTIAL).
 FT EXTRACELLULAR (POTENTIAL).


```
Db 265 GVTDEI-YQGPIVLLNIFLENIVRLMTKLRASTSETIQYKAVKATVLLP 323
QY 487 GVTWGLAIFTP-----LGLSTVYIFALFNSLOGVFCICWFTIL 524
Db 324 GITYMLFFVTGGEISRIYFIYFNSLQSFQGFVSVFYCF 366

RESULT 15
ID CRP2_XENLA STANDARD; PRT; 413 AA.
AC 042603;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Corticotropin releasing factor receptor 2 precursor (CRF-R 2) (CRF2)
DE (Corticotropin-releasing hormone receptor 2) (CRH-R 2).
GN CRF2.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain, and Heart;
RX MEDLINE=97465573; PubMed=9326293;
RA Dautzenberg F.M., Dietrich K., Palchaudhuri M.R., Spiess J.;
RT "Identification of two corticotropin-releasing factor receptors from
RT Xenopus laevis with high ligand selectivity: unusual pharmacology of
RT the type 1 receptor."
RL J. Neurochem. 69:1640-1649(1997).
CC -!- FUNCTION: THIS IS A RECEPTOR FOR CORTICOTROPIN RELEASING FACTOR.
CC SHOWS HIGH-AFFINITY BINDING FOR UROTENSIN I. THE ACTIVITY OF THIS
CC RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE ADENYL
CC CYCLASE (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: Y14037; CAA74364.1; -.
CC GCRDB: GCR_2578; -.
CC InterPro: IPR000832; GPCR_secretin.
CC InterPro: IPR001879; hormn_receptor.
CC Pfam: PF00002; 7tm_2; 1.
CC Pfam: PF02793; HRM; 1.
CC PRINTS: PR00249; GPCRSECRETIN.
CC SMART: SM00008; Hormr; 1.
CC PROSITE: PS00649; G_PROTEIN_RECEP_F2.1; 1.
CC PROSITE: PS00650; G_PROTEIN_RECEP_F2.2; 1.
CC PROSITE: PS00650; G_PROTEIN_RECEP_F2.3; 1.
CC PROSITE: PS0227; G_PROTEIN_RECEP_F2.4; 1.
CC PROSITE: PS0261; G_PROTEIN_RECEP_F2.4; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 ?
FT CHAIN ? 413 CORTICOTROPIN RELEASING FACTOR RECEPTOR
FT FT ? 2.
FT DOMAIN ? 120 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 121 141 1 (POTENTIAL).
FT DOMAIN 142 150 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 151 170 2 (POTENTIAL).
FT DOMAIN 171 187 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 188 211 3 (POTENTIAL).
FT DOMAIN 212 225 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 226 247 4 (POTENTIAL).
FT DOMAIN 248 266 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 267 289 5 (POTENTIAL).
```

```
FT DOMAIN 290 312 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 313 332 6 (POTENTIAL).
FT DOMAIN 333 347 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 348 367 7 (POTENTIAL).
FT DOMAIN 368 413 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 16 16 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 77 77 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 89 89 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 97 97 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 413 AA; 48458 MW; DAD422F0A96C4626 CRC64;

Query Match 5.6%; Score 161.5; DB 1; Length 413;
Best Local Similarity 25.6%; Pred. No. 8.4e-05;
Matches 72; Conservative 45; Mismatches 131; Indels 33; Gaps 12;

QY 258 PTLDQS---TVH--ILTRISAGCGVSMIFLAFTIILYAFRLSRERFKSEDAPKIHVAL 312
Db 103 PILDNRKRYALHYKIALIINYLGHCISILALVIAFLFLCLRSIRCLRN-----IHNL 157
QY 313 GGSFLFLNLA-FLVNVGSGSKGSDAACWAR--GAVFHYFLICAFWMLGFAFHLYLLAVR 369
Db 158 ITTFLRNIMWFLQMDIHNIHESNEVWCRCITTIYNYFVVTNFFWFEVCYLHTAIVM 217
QY 370 VNTYFGHYFLKLSLVGWGLPALMVIGTGSANSYGLYTIIRDRENTSLELCWFREGTMY 429
Db 218 TVSTDKLRKQVFL-FIGWCIPSPIIIVTWAICKLF-----YEN---EQCWIGKEPKY 265
QY 430 ALYITVHGIFLITFLFGMVVLAIVYWKFTLSRATAYKRGKKNKKVLTLLGLSSLVGVT 489
Db 266 IDYI-YQGRVTLVLLINFLVFLNIVRLMTKLRASTSETIQYKAVKATVLLP 324
QY 490 WGLAIFTPLGLSTV--YIFALFN---SLQGVFICCMFTIL 524
Db 325 YMLFEVNP-GEDDYSQIVFIYFNSFLQSFQGFVSVFYCF 364
```

Search completed: June 11, 2002, 22:30:41
Job time: 430 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 11, 2002, 22:24:36 ; Search time 68.53 Seconds
(without alignments)
1385.878 Million cell updates/sec

Title: US-09-733-387-44

Perfect score: 2878

Sequence: 1 MATPRGLGALLLLLLLPTSG.....STTVSSSTARLQAHASQAE 549

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_19:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phage:*

10: sp_plant:*

11: sp_rhodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_rvirus:*

16: sp_bacteriap:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	643.5	22.4	687	11 Q9QZT2	Q9qzt2 mus musculus
2	610.5	21.2	1014	4 Q00406	Q00406 homo sapien
3	603.5	21.0	687	4 Q95966	Q95966 homo sapien
4	596.5	20.7	693	4 Q96HB4	Q96hb4 homo sapien
5	594.5	20.7	693	4 Q9Y653	Q9y653 homo sapien
6	554.5	19.3	512	4 Q96JW0	Q96jw0 homo sapien
7	555.5	19.3	541	4 Q9Y3K0	Q9y3k0 homo sapien
8	387	13.4	872	4 Q94867	Q94867 homo sapien
9	383.5	13.3	1231	6 Q97819	Q97819 bos taurus
10	383.5	13.3	1274	6 Q97820	Q97820 bos taurus
11	383.5	13.3	1299	6 Q97825	Q97825 bos taurus
12	383.5	13.3	1342	6 Q97826	Q97826 bos taurus
13	383.5	13.3	1503	6 Q97818	Q97818 bos taurus
14	383.5	13.3	1527	11 Q88927	Q88927 rattus norv
15	383.5	13.3	1571	6 Q97824	Q97824 bos taurus
16	382.5	13.3	823	4 Q9UHX3	Q9uhx3 homo sapien

17	382.5	13.3	1240	4 Q9HAR2	Q9har2 homo sapien
18	379	13.2	1240	6 Q97822	Q97822 bos taurus
19	379	13.2	1283	6 Q97823	Q97823 bos taurus
20	379	13.2	1308	6 Q97828	Q97828 bos taurus
21	379	13.2	1351	6 Q97829	Q97829 bos taurus
22	379	13.2	1512	6 Q97821	Q97821 bos taurus
23	379	13.2	1550	11 Q9Z173	Q9z173 rattus norv
24	379	13.2	1580	6 Q97827	Q97827 bos taurus
25	371.5	12.9	839	4 Q96IE7	Q96ie7 homo sapien
26	371.5	12.9	1469	4 Q9HAR3	Q9har3 homo sapien
27	371.5	12.9	1474	4 Q94910	Q94910 homo sapien
28	370.5	12.9	1467	6 Q97830	Q97830 bos taurus
29	370.5	12.9	1472	6 Q97831	Q97831 bos taurus
30	365.5	12.7	1515	11 Q88917	Q88917 rattus norv
31	361.5	12.6	1021	4 Q94882	Q94882 homo sapien
32	359.5	12.5	1354	6 Q97806	Q97806 bos taurus
33	359.5	12.5	1397	6 Q97807	Q97807 bos taurus
34	359.5	12.5	1420	6 Q97814	Q97814 bos taurus
35	359.5	12.5	1463	6 Q97815	Q97815 bos taurus
36	355.5	12.4	1478	11 Q88923	Q88923 rattus norv
37	355	12.3	1123	4 Q9UKY5	Q9uky5 homo sapien
38	355	12.3	1177	4 Q9UKY6	Q9uky6 homo sapien
39	355	12.3	1403	4 Q95490	Q95490 homo sapien
40	353	12.3	1341	6 Q97802	Q97802 bos taurus
41	353	12.3	1384	6 Q97803	Q97803 bos taurus
42	353	12.3	1407	6 Q97810	Q97810 bos taurus
43	353	12.3	1450	6 Q97811	Q97811 bos taurus
44	351	12.2	606	4 Q9HBW9	Q9hbW9 homo sapien
45	351	12.2	1487	11 Q9Z174	Q9z174 rattus norv

ALIGNMENTS

RESULT 1

Q9QZT2	ID	Q9QZT2	PRELIMINARY;	PRT;	687 AA.
AC	Q9QZT2:				
DT	01-MAY-2000	(TREMBLrel. 13, Created)			
DT	01-MAY-2000	(TREMBLrel. 13, Last sequence update)			
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)			
DE	SERPENTINE RECEPTOR.				
GN	GPR56 OR CYT28.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX	NCBI_TaxID=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=C57BL/6; TISSUE=LIVER;				
RA	Phillips R.L., Ernst R.E., Dosisl M., Wesley C.K., Moore K.A.,				
RA	Kingsley P.D., Sykes S., Palis J., Lemischka I.R.;				
RT	"Identification of novel hematopoietic stem cell regulatory genes."				
RL	Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AF166382; AAF0617.1;				
DR	MGD; MGI:1340051; Gpr56				
DR	InterPro; IPR000873; AMP-bind.				
DR	InterPro; IPR000832; GPCR_secretin.				
DR	InterPro; IPR000203; PKD_cys_rich.				
DR	Pfam; PF00002; 7tm.2; 1.				
DR	Pfam; PF01825; GPS; 1.				
DR	PRINTS; PR00249; GPCRSECRETIN.				
DR	SMART; SM00303; GPS; 1.				
DR	PROSITE; PS00455; AMP_BINDING; UNKNOWN_1.				
DR	PROSITE; PS0261; G_PROTEIN_RECEPTOR_F2_4; 1.				
KW	Receptor.				
SQ	SEQUENCE	687 AA;	77271 MW;	B5315D70AF443809	CRC64;

Query Match 22.4%; Score 643.5; DB 11; Length 687;
Best Local Similarity 31.1%; Pred. No. 1.2e-47;
Matches 168; Conservative 89; Mismatches 219; Indels 65; Gaps 17;

QY 307 KIHVALGSGIFLLNLAFLVNWSGSGSDAACWARGAVHYFLLCATFWMGLEAFHLYLL 366
 I:-I : : :::: :|||: |:-| : |||: |:-| : |||: |:-| : |||: |:-| : |||:
 Db 438 KYVMMLLAVFLDTSFLLSEPVALTSGEAGCRASAIPLHFHSLLTCLSMWGLEGYNLYRL 497

QY 367 AYRVNTYFGHYFLKLSIVGWGLPALMV--IGTGSANSYG---LYTIRDRENTSLELCW 421
 I:-I : : |||: |:-| : |||: |:-| : |||: |:-| : |||: |:-| : |||: |:-| : |||:
 Db 498 VVEVGTVPGLTKLSAMGWGPFLVTILVALVDVNGPIILLAVHTPEGVYPSMCW 557

QY 422 FREGTMYALITYHGHYFLTFTFGWMVLALVWKIFTLSRATAVKRGKNRKVVLTLLG 481
 I:-I : : |||: |:-| : |||: |:-| : |||: |:-| : |||: |:-| : |||: |:-| : |||:
 Db 558 IRDSLVS---YITNLGLFSLVLFNMAMLAITVQILRLPHT-----QKSWHLTLTG 608

QY 482 LSLVGVTVGLAIFT---PLGISTVIIFALPNSLOGVFICWFCTILYLSPQS-----TTV 533
 I:-I : : |||: |:-| : |||: |:-| : |||: |:-| : |||: |:-| : |||: |:-| : |||:
 Db 609 LSLVIGLPMAIFFSFASGTQLVWLIFSITSOGFLIFTWMSRMLOARQGPSPLKS 668

QY 534 SSSTARL 540
 I:-I : : |||: |:-| : |||: |:-| : |||: |:-| : |||: |:-| : |||: |:-| : |||:
 Db 669 NSDSARL 675

RESULT 4
 Q96HB4 PRELIMINARY; PRT; 693 AA.
 ID AC Q96HB4;
 DT 01-DEC-2001 (TreeBLrel. 19, Created)
 DT 01-DEC-2001 (TreeBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TreeBLrel. 19, Last annotation update)
 G PROTEIN-COUPLED RECEPTOR 56.
 OS Homo sapiens (Human).
 DE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=PLACENTA, AND CHORIOCARCINOMA;
 RA Strausberg R.;
 RL Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
 RW EMBL; BC008770; AAH08770.1; --
 KW Receptor.
 SQ SEQUENCE 693 AA; 77746 MW; 5C9FA8194D1B5C50 CRC64;

Query Match 20.7%; Score 596.5; DB 4; Length 693;
 Best Local Similarity 34.7%; Pred. No. 1.5e-43;
 Matches 149; Conservative 66; Mismatches 181; Indels 33; Gaps

QY 131 VRLPKSLRPSIPGNRSVVRVAVTILDIGPCTLFKGPRGLGDGSGVLNNRLVGLSYCOMH 190
 I:-I : : |||: |:-| : |||: |:-| : |||: |:-| : |||: |:-| : |||: |:-| : |||:
 Db 267 VLLPTFLQTQKRGSGEAERKLLLDVDFSSQAUFQK----NSSHVJGEKVILGIVONTK 321

QY 191 VTKLAEPLIEIVFSHORPPNNMTTCVFW--DVTKGITGDWSEGGSTEVRPEGIVCCCHD 248
 I:-I : : |||: |:-| : |||: |:-| : |||: |:-| : |||: |:-| : |||: |:-| : |||:
 Db 322 VANLTEPVVLTFOHQLOKPNVTLQCWFVEDPTLSSPGHWSSAGCETVRETQOTSCFCNH 381

QY 249 LPFFALLLRPTLDQSTV--HILIRISQAGCGYSMFIAFTIILYAFLR--LSRERFKSED 304
 I:-I : : |||: |:-| : |||: |:-| : |||: |:-| : |||: |:-| : |||: |:-| : |||:
 Db 382 LTAFYAWLYSVSSVDVAKHKYLSLLSYVCWCWSALACLVTIYAALCSRPVLPCCRKPXY 441

QY 305 APKIHVALGGSFLFLNLAFILVNWSGSGSKGSDAACWARGAVHYFLLCATFWMGLEAFHLY 364
 I:-I : : |||: |:-| : |||: |:-| : |||: |:-| : |||: |:-| : |||: |:-| : |||:
 Db 442 TKVHMNLLAVFLDTSFLLSEPVALTSGEAGCRASAIFLHFSLTCLSMWGLEGYNLY 501

QY 365 LLAVRVNTYFGHYFLKLSIVGWGLPALMV--IGTGSANSYG---LYTIRDRENTSLEL 419
 I:-I : : |||: |:-| : |||: |:-| : |||: |:-| : |||: |:-| : |||: |:-| : |||:
 Db 502 RLVEVFGTYTPGYLLKLSAMGWGFIFLVTILVALVDVNGPIILLAVHTPEGVYPSM 561

QY 420 CWFREGITMYALITYHGHYFLTFTFGWMVLALVWKIFTLSRATAVKRGKNRKVVLT 479
 I:-I : : |||: |:-| : |||: |:-| : |||: |:-| : |||: |:-| : |||: |:-| : |||:
 Db 562 CWRDLSVS---YITNLGLFSLVLFNMAMLAITVQILRLPHT-----QKWSHVLTL 612

QY 480 LGSLSLVGTVTWGLAIFT---PLGISVTYIIFALPNSLOGVFICWFCTILYLSPQS-----T 531

[illegible]

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99148828; PubMed=10025961;
RA Matsushita H., Lelanova V.G., Ushkaryov Y.A.;
RT "The latrophilin family: multiply spliced G protein-coupled receptors
with differential tissue distribution.";
RL FEBS Lett. 443:348-352(1999).
DR EMBL; AF111086; AAD05322.1; .
DR InterPro; IPR000922; Gal_lectin.
DR InterPro; IPR000832; GPCR_secretin.
DR InterPro; IPR001879; hormn_receptor.
DR InterPro; IPR003334; Latrophilin.
DR InterPro; IPR003112; Olfac_like.
DR InterPro; IPR00203; PKD_cys_rich.
DR Pfam; PF00002; 7tm_2; 1.
DR Pfam; PF02140; Gal_lectin; 1.
DR Pfam; PF01825; GPS; 1.
DR Pfam; PF02793; HRM; 1.
DR Pfam; PF02191; OLF; 1.
DR PRINTS; PR00249; GPCRSECRETIN.
DR ProDom; PD005612; Gal_lectin; 1.
DR SMART; SM00303; GPS; 1.
DR SMART; SM00008; Hormr; 1.
DR SMART; SM00284; OLF; 1.
DR PROSITE; PS0227; G_PROTEIN_RECEP_F2_3; 1.
DR PROSITE; PS0261; G_PROTEIN_RECEP_F2_4; 1.
DR PROSITE; PS0228; SUEL_LECTIN; 1.
SQ SEQUENCE 1231 AA; 138200 MW; 1258A8A05D64325D CRC64;

Query Match 13.3%; Score 383.5; DB 6; Length 1231;
Best Local Similarity 27.6%; Pred. No. 1.3e-24;
Matches 135; Conservative 77; Mismatches 182; Indels 95; Gaps 25;
QY 77 HLMKEGLTKVNTFFLKALVONLST-NTAEDFYFSLEPS-----QVPRQVMKDEKPPDR 130
DB 683 NLLKTDIVRE-NTDNQLEVARLSTEGNLEDKFPENTGHSTGTLQLSANTLKN----- 735
QY 131 VRLPKSLFRSLPGNRSVRLAVTIL-DIGPGTLFGPRLGLG-----DGSGLVNNRLVG 183
DB 736 -----GRNGEIRVAFVLYNGLPYLSTENASMKLGTTEAMSTNHSVIVNSPVIT 783
QY 184 LSVGMQHVTK--LAEPLEIVESH-QRPPNNMTLCVFDVTKGT-TGDSSEGCs-TEVR 238
DB 784 AAINKEFSNKVYLPADPVVFTVKHKQSEENFNPCSFWSYKRTMTGWSQGCRLLTTN 843
QY 239 PEGTVCCDHLTFFALLR--PTLDQSTVH--ILTRISQAGCGVSMIFLAFTIILYAFLR 294
DB 844 KTHHTCSNHLTNFAVMAHVEVKHSDAHDLLDVTWVIGLLSLVCLLICITFCFF- 902
QY 295 LSREKFSKEDAPKIHVALGSLFLNLAFVNVGSGKSGSDAACWARGAVHYFLCAFT 354
DB 903 -----RGLQSDRNTIHKNCISLFAELLFL--IGINTDQPIACAVFAALLHFFFLAFT 956
QY 355 WMGLEAFHLYLLAVRVNTYFGH--YFLKLSLVGWLGPALMVGITGSAN--SYGLYTIRD 410
DB 957 WMFLEGVQLYIMLVEVFSEHSRRKIFY--LVGYGMPALIVASAAVDYSYG----- 1007
QY 411 RENRTSLELCWPREGTTMYALYIVVHGYFLITF-----LFGMVVALVVMKIFLISR 462
DB 1008 -----TDKVCWLRD-----YFIWSFIGPATLIIMNLVIFLGIALYKMF----H 1048
QY 463 ATAV--KERK--KNRKK-VLTLLGLSSLVGVTW--GLAIFTPGLSTYVIFALNSLOGV 515
DB 1049 HTAILKPESGCCDNKSWVIGAIALLCLLGLTWAFGLMYINESTVIMAYLFTIFNSLOGM 1108
QY 516 FICCWFTIL 524
|| : : ||

Db 1109 FIFIFHCVL 1117
RESULT 10
ID O97820 PRELIMINARY; PRT: 1274 AA.
AC O97820;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE LATROPHILIN 3 SPLICE VARIANT ABAB.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99148828; PubMed=10025961;
RA Matsushita H., Lelanova V.G., Ushkaryov Y.A.;
RT "The latrophilin family: multiply spliced G protein-coupled receptors
with differential tissue distribution.";
RL FEBS Lett. 443:348-352(1999).
DR EMBL; AF111087; AAD05323.1; .
DR InterPro; IPR000922; Gal_lectin.
DR InterPro; IPR000832; GPCR_secretin.
DR InterPro; IPR001879; hormn_receptor.
DR InterPro; IPR003334; Latrophilin.
DR InterPro; IPR003112; Olfac_like.
DR InterPro; IPR00203; PKD_cys_rich.
DR Pfam; PF00002; 7tm_2; 1.
DR Pfam; PF02140; Gal_lectin; 1.
DR Pfam; PF01825; GPS; 1.
DR Pfam; PF02793; HRM; 1.
DR Pfam; PF02354; Latrophilin; 1.
DR Pfam; PF02191; OLF; 1.
DR PRINTS; PR00249; GPCRSECRETIN.
DR ProDom; PD005612; Gal_lectin; 1.
DR SMART; SM00303; GPS; 1.
DR SMART; SM00008; Hormr; 1.
DR SMART; SM00284; OLF; 1.
DR PROSITE; PS0227; G_PROTEIN_RECEP_F2_3; 1.
DR PROSITE; PS0261; G_PROTEIN_RECEP_F2_4; 1.
DR PROSITE; PS0228; SUEL_LECTIN; 1.
SQ SEQUENCE 1274 AA; 142748 MW; 147C1B5FC160037F CRC64;
Query Match 13.3%; Score 383.5; DB 6; Length 1274;
Best Local Similarity 27.6%; Pred. No. 1.4e-24;
Matches 135; Conservative 77; Mismatches 182; Indels 95; Gaps 25;
QY 77 HLMKEGLTKVNTFFLKALVONLST-NTAEDFYFSLEPS-----QVPRQVMKDEKPPDR 130
DB 683 NLLKTDIVRE-NTDNQLEVARLSTEGNLEDKFPENTGHSTGTLQLSANTLKN----- 735
QY 131 VRLPKSLFRSLPGNRSVRLAVTIL-DIGPGTLFGPRLGLG-----DGSGLVNNRLVG 183
DB 736 -----GRNGEIRVAFVLYNGLPYLSTENASMKLGTTEAMSTNHSVIVNSPVIT 783
QY 184 LSVGMQHVTK--LAEPLEIVESH-QRPPNNMTLCVFDVTKGT-TGDSSEGCs-TEVR 238
DB 784 AAINKEFSNKVYLPADPVVFTVKHKQSEENFNPCSFWSYKRTMTGWSQGCRLLTTN 843
QY 239 PEGTVCCDHLTFFALLR--PTLDQSTVH--ILTRISQAGCGVSMIFLAFTIILYAFLR 294
DB 844 KTHHTCSNHLTNFAVMAHVEVKHSDAHDLLDVTWVIGLLSLVCLLICITFCFF- 902
QY 295 LSREKFSKEDAPKIHVALGSLFLNLAFVNVGSGKSGSDAACWARGAVHYFLCAFT 354
DB 903 -----RGLQSDRNTIHKNCISLFAELLFL--IGINTDQPIACAVFAALLHFFFLAFT 956
QY 355 WMGLEAFHLYLLAVRVNTYFGH--YFLKLSLVGWLGPALMVGITGSAN--SYGLYTIRD 410
|| : : ||

```
Db 957 WMFEGVQLYIMLVEFESEHSRRKYFY---LVGYGMPALIVAVSAADVRSYG----- 1007
QY 411 RENRTSLELCWREGTMYALYITVHGXYLTF-----LFGMVVLALVVKWKTLSR 462
Db 1008 -----TDKVCWLRLD-----YFWSFGPATLIIMLVIFGLIYKMF---H 1048
QY 463 ATAV--KRRK--VLTLLGSSLVGVTV--GLAFTPLGLSTVYIFALFNSLQGV 515
Db 1049 HTAILKPESGCLDNIKSWIGAIALLCGLTWAFGLMYINESTVIMAYLFTIFNSLQGM 1108
QY 516 FICCWFTIL 524
Db 1109 FIFIFHCVL 1117

RESULT 11
O97825 PRELIMINARY; PRT; 1299 AA.
AC O97825;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-OCT-2001 (Tremblrel. 18, Last annotation update)
DE LATROPHILIN 3 SPLICED VARIANT BBAG.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99148828; PubMed=10025961;
RA Matsushita H., Lelianova V.G., Ushkaryov Y.A.;
RT "The latrophilin family: multiply spliced G protein-coupled receptors
with differential tissue distribution.";
RL FEBS Lett. 443:348-352(1999).
DR EMBL; AF111092; AAD05328.1; -
DR InterPro; IPR000922; Gal_lectin.
DR InterPro; IPR000832; GPCR_secretin.
DR InterPro; IPR001879; hormn_receptor.
DR InterPro; IPR003334; Latrophilin.
DR InterPro; IPR003112; Olfac_like.
DR InterPro; IPR000203; PKD_cys_rich.
DR Pfam; PF00002; 7tm2; 1.
DR Pfam; PF02140; Gal_lectin; 1.
DR Pfam; PF01825; GPS; 1.
DR Pfam; PF02793; HRM; 1.
DR Pfam; PF02354; Latrophilin; 1.
DR Pfam; PF02191; OLF; 1.
DR PRINTS; PR00249; GPCRSECRETIN.
DR SMART; SM00303; GPS; 1.
DR SMART; SM00284; OLF; 1.
DR PROSITE; PS50227; G_PROTEIN_RECEP_F2_3; 1.
DR PROSITE; PS50261; G_PROTEIN_RECEP_F2_4; 1.
DR PROSITE; PS50228; SUEL_LECTIN; 1.
SQ SEQUENCE 1299 AA; 145109 MW; 9D83D9258D7FA821 CRC64;

Query Match 13.3%; Score 383.5; DB 6; Length 1299;
Best Local Similarity 27.6%; Pred. No. 1.4e-24;
Matches 135; Conservative 77; Mismatches 182; Indels 95; Gaps 25;

QY 77 HLMKEGLTQKVTFFFLKALYQNLST-NTAEDFYFSLEPS-----QVPRQVMKDEKPPDR 130
Db 751 NLLKTDIVRE-NTNDTIQLEVARLSTEGCNLEDLKFPENTGHGSTIQLSANTLKQN----- 803
QY 131 VRLPKSLFRSLPCNRSVRLAVTIL-DIGPGTLFKGPRGLG-----DGSGVLNRLVG 183
Db 804 -----GRNGEIRVAFVLYNNLGPYLSTENASMKLGTTEAMSTNHSVIVNSPVIT 851
QY 184 LSYGQMHVTK--LAEPLEIVFSH-QRPPPNMTLTCVFDVTKGT-TGDWSSEGCs-TEVR 238
::: |::|:: |::|:: |::|:: |::|:: |::|:: |::|:: |::|:: |::|:: |::|::
```

```
Db 852 AAINKEFSNKVYLADPVVFTVKHIKQSEENFNPNCSFWSYSKRTMTGYWSTQGCRLTTN 911
QY 239 PEGTVCCDHLTFEALLR--PTLDQSTVH--ILTRISOAGCGVSMIFLAFTIILYAFLR 294
Db 912 KTHITCSNHLTNFAVLMAHVEVKHSDAHDLLDLDVTWVGILLSLVCLICITFCFF- 970
QY 295 LSRRFKSEDAKTHVALGGSLFLNLIAFLVNVGSGSGSDAACWARGAVHYFLCAFT 354
Db 971 ---RGLQSDRNTIHKNLICISLFAELLFL--IGINRTDQPIACAVEAALLHFFFLAAFT 1024
QY 355 WMGLEAFHLYLLAVRVNTYFGH--YFKLSLVGWLGPALMVIGTGSAN--SYGLYIRD 410
Db 1025 WMFEGVQLYIMLVEFESEHSRRKYFY---LVGYGMPALIVAVSAADVRSYG----- 1075
QY 411 RENRTSLELCWREGTMYALYITVHGXYLTF-----LFGMVVLALVVKWKTLSR 462
Db 1076 -----TDKVCWLRLD-----YFWSFGPATLIIMLVIFGLIYKMF---H 1116
QY 463 ATAV--KRRK--VLTLLGSSLVGVTV--GLAFTPLGLSTVYIFALFNSLQGV 515
Db 1117 HTAILKPESGCLDNIKSWIGAIALLCGLTWAFGLMYINESTVIMAYLFTIFNSLQGM 1176
QY 516 FICCWFTIL 524
Db 1177 FIFIFHCVL 1185

RESULT 12
O97826 PRELIMINARY; PRT; 1342 AA.
AC O97826;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE LATROPHILIN 3 SPLICED VARIANT BBAG.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99148828; PubMed=10025961;
RA Matsushita H., Lelianova V.G., Ushkaryov Y.A.;
RT "The latrophilin family: multiply spliced G protein-coupled receptors
with differential tissue distribution.";
RL FEBS Lett. 443:348-352(1999).
DR EMBL; AF111093; AAD05329.1; -
DR InterPro; IPR000922; Gal_lectin.
DR InterPro; IPR000832; GPCR_secretin.
DR InterPro; IPR001879; hormn_receptor.
DR InterPro; IPR003334; Latrophilin.
DR InterPro; IPR003112; Olfac_like.
DR InterPro; IPR000203; PKD_cys_rich.
DR Pfam; PF00002; 7tm2; 1.
DR Pfam; PF02140; Gal_lectin; 1.
DR Pfam; PF01825; GPS; 1.
DR Pfam; PF02793; HRM; 1.
DR Pfam; PF02354; Latrophilin; 1.
DR Pfam; PF02191; OLF; 1.
DR PRINTS; PR00249; GPCRSECRETIN.
DR PRODOM; PD005612; Gal_lectin; 1.
DR SMART; SM00303; GPS; 1.
DR SMART; SM00008; Hormr; 1.
DR SMART; SM00284; OLF; 1.
DR PROSITE; PS50227; G_PROTEIN_RECEP_F2_3; 1.
DR PROSITE; PS50261; G_PROTEIN_RECEP_F2_4; 1.
DR PROSITE; PS50228; SUEL_LECTIN; 1.
SQ SEQUENCE 1342 AA; 149658 MW; 51FBC79F37FF5B4B CRC64;

Query Match 13.3%; Score 383.5; DB 6; Length 1342;
Best Local Similarity 27.6%; Pred. No. 1.5e-24;
```

```
Matches 135; Conservative 77; Mismatches 182; Indels 95; Gaps 25;

QY 77 HLMKEGLTKVNTPELKALVQLNST-NTAEDFYFSLEPS-----QVPRQVMKDEKPPDR 130
D 751 NLLKTDIVRE-NTDNIQLEVARLSTEGNLEDKFPENTGHSTIQLSANTLKN-----803
QY 131 VRLPKSLFRSLPGNRSVRLAVTIL-DIGPGTLFGKPRGLG-----DGSGLVNLRLVG 183
D 804 -----GRNGEIRVAFVLYNNGPYLSTENASMKLGTEAMSTNHSVIVNSPVIT 851
QY 184 LSVGQMHVTK--LAEPLEIVFESH--ORPPNMTLTCVFDVTKGT-TGDWSSEGCs-TEVR 238
D 852 AAINKEFSNKVYLADPVVFTVKHIKQSEENFNPCSFWSYKRTMTGYWSTQGCRLLTN 911
QY 239 PEGTVCCDHLTFPALLR--PTLDQSTVH--ILTRISQAGCGVSMIFLAFTIILYAFRL 294
D 912 KTHTTCSNHLTNFAVLAHVEVKHSDAVHDLDDVITWVGILLSVCLLICITFCFF- 970
QY 295 LSRRFKSEDAPKIHVALGSLFLNLAFVNVGSGSDAACWARGAVHYFLLCFT 354
D 971 -----RGLQSDRNTIHKNCISLFAELLFL--IGINRTDQPIACAVFAALLHFFFLAAET 1024
QY 355 WMGLEAFHLYLLAVRVNTYFGH--YFLKLSLVGWLPLAMVIGTGSAN--SYGLYTIRD 410
D 1025 WMFLEGVQLYIMLVEVESEHSRRKIFY---LVGYMPALIVAVSAADVRSIG-----1075
QY 411 RENRTSLELCWFREGTMYALYITVHGFLITF-----LFGMVVLALVVMKIFTLR 462
D 1076 -----TDKVCWLRD-----YFIWSFGPATLIIMLVIFGLIALYKMF---H 1116
QY 463 ATAV--KERG--KNRKK-VLTLLGLSLVGVTV--GLAIFTPLGLSTVYIFALFNSLOGV 515
D 1117 HTAILKPESGCLDNKISWVIGAILLGLLTWAFGLMYINESTVIMAYLFTIFNSLOGM 1176
QY 516 FICCWFTIL 524
D 1177 FIFIFHCVL 1185

RESULT 13
ID 097818 PRELIMINARY; PRT; 1503 AA.
AC 097818;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE LATROPHILIN 3 SPLICE VARIANT ABAB.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99148828; PubMed=10025961;
RA Matsushita H., Lelianaova V.G., Ushkaryov Y.A.;
RT "The latrophilin family: multiply spliced G protein-coupled receptors
RT with differential tissue distribution."
RL FBS Lett. 443:348-352(1999).
DR EMBL; AF111085; AAD05321.1;
DR InterPro; IPR000922; Gal.lectin.
DR InterPro; IPR000832; GPCR_secretin.
DR InterPro; IPR001879; horrm_receptor.
DR InterPro; IPR003334; Latrophilin.
DR InterPro; IPR003112; Olfac_like.
DR InterPro; IPR000203; PKD_cys_rich.
DR Pfam; PF00002; 7tm.2; 1.
DR Pfam; PF02140; Gal.lectin; 1.
DR Pfam; PF01925; GPS; 1.
DR Pfam; PF02793; HRM; 1.
DR Pfam; PF02354; Latrophilin; 1.
DR Pfam; PF02191; OLF; 1.
DR PRINTS; PR00249; GPCRSECRETIN.
```

```
ProDom; PD005612; Gal.lectin; 1.
DR SMART; SM00303; GPS; 1.
DR SMART; SM00008; Horrm; 1.
DR SMART; SM00284; OLF; 1.
DR PROSITE; PS50227; G_PROTEIN_RECIP_F2_3; 1.
DR PROSITE; PS50261; G_PROTEIN_RECIP_F2_4; 1.
DR PROSITE; PS50228; SUBEL_LECTIN; 1.
SQ SEQUENCE 1503 AA; 167991 MW; F4C8345D7F988218 CRC64;

Query Match 13.3%; Score 383.5; DB 6; Length 1503;
Best Local Similarity 27.6%; Pred. No. 1.7e-24;
Matches 135; Conservative 77; Mismatches 182; Indels 95; Gaps 25;

QY 77 HLMKEGLTKVNTPELKALVQLNST-NTAEDFYFSLEPS-----QVPRQVMKDEKPPDR 130
D 683 NLLKTDIVRE-NTDNIQLEVARLSTEGNLEDKFPENTGHSTIQLSANTLKN-----735
QY 131 VRLPKSLFRSLPGNRSVRLAVTIL-DIGPGTLFGKPRGLG-----DGSGLVNLRLVG 183
D 736 -----GRNGEIRVAFVLYNNGPYLSTENASMKLGTEAMSTNHSVIVNSPVIT 783
QY 184 LSVGQMHVTK--LAEPLEIVFESH--ORPPNMTLTCVFDVTKGT-TGDWSSEGCs-TEVR 238
D 784 AAINKEFSNKVYLADPVVFTVKHIKQSEENFNPCSFWSYKRTMTGYWSTQGCRLLTN 843
QY 239 PEGTVCCDHLTFPALLR--PTLDQSTVH--ILTRISQAGCGVSMIFLAFTIILYAFRL 294
D 844 KTHTTCSNHLTNFAVLAHVEVKHSDAVHDLDDVITWVGILLSVCLLICITFCFF- 902
QY 295 LSRRFKSEDAPKIHVALGSLFLNLAFVNVGSGSDAACWARGAVHYFLLCFT 354
D 903 -----RGLQSDRNTIHKNCISLFAELLFL--IGINRTDQPIACAVFAALLHFFFLAAET 956
QY 355 WMGLEAFHLYLLAVRVNTYFGH--YFLKLSLVGWLPLAMVIGTGSAN--SYGLYTIRD 410
D 957 WMFLEGVQLYIMLVEVESEHSRRKIFY---LVGYMPALIVAVSAADVRSIG-----1007
QY 411 RENRTSLELCWFREGTMYALYITVHGFLITF-----LFGMVVLALVVMKIFTLR 462
D 1008 -----TDKVCWLRD-----YFIWSFGPATLIIMLVIFGLIALYKMF---H 1048
QY 463 ATAV--KERG--KNRKK-VLTLLGLSLVGVTV--GLAIFTPLGLSTVYIFALFNSLOGV 515
D 1049 HTAILKPESGCLDNKISWVIGAILLGLLTWAFGLMYINESTVIMAYLFTIFNSLOGM 1108
QY 516 FICCWFTIL 524
D 1109 FIFIFHCVL 1117

RESULT 14
ID 088927 PRELIMINARY; PRT; 1527 AA.
AC 088927; O88924; O88928; O88929; O88926; O88925;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE CALCIUM-INDEPENDENT ALPHA-LATROTOXIN RECEPTOR 3 PRECURSOR (LATROPHILIN
DE 3) (LRP3) (CIRL) (CL3) (LPH3).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99047651; PubMed=9830014;
RA Sugita S., Ichtchenko K., Khvotchev M., Suedhof T.C.;
RT "alpha-Latrotoxin receptor CIRL/Latrophilin 1 (CL1) defines an unusual
RT family of ubiquitous G-protein-linked receptors. G-protein coupling
RT not required for triggering exocytosis."
RL J. Biol. Chem. 273:32715-32724(1998).
CC -!- FUNCTION: PERFORMS A GENERAL AND UBIQUITOUS FUNCTION AS G-PROTEIN-
```


DR InterPro; IPR000832; GPCR_secretin.
DR InterPro; IPR001879; hormn_receptor.
DR InterPro; IPR003334; latrophilin.
DR InterPro; IPR003112; Olfac_like.
DR InterPro; IPR000203; PKD_cys_rich.
DR Pfam; PF00002; 7tm_2; 1.
DR Pfam; PF02140; Gal_Lectin; 1.
DR Pfam; PF01825; GPS; 1.
DR Pfam; PF02793; HRM; 1.
DR Pfam; PF02354; Latrophilin; 1.
DR Pfam; PF02191; OLF; 1.
DR PRINTS; PR00249; GPCRSECRETIN.
DR PRODom; PD005612; Gal_lectin; 1.
DR SMART; SM00303; GPS; 1.
DR SMART; SM00008; Hormr; 1.
DR SMART; SM00284; OLF; 1.
DR PROSITE; PS02027; G_PROTEIN_RECEP_F2_3; 1.
DR PROSITE; PS02061; G_PROTEIN_RECEP_F2_4; 1.
DR PROSITE; PS02028; SUEL_LECTIN; 1.
SQ SEQUENCE 1571 AA; 174901 MW; 2185D5ADCB5F8607 CRC64;

Query Match 13.3%; Score 383.5; DB 6; Length 1571;
Best Local Similarity 27.6%; Pred. No. 1.8e-24;
Matches 135; Conservative 77; Mismatches 182; Indels 95; Gaps 25;

Qy 77 HLMKEGLTQKYNTPFLKALVONLST-NTAEDFYFSLEPS-----QVPRQVWKDEKPPDR 130
Db 751 NLLKTDIVRE-NTDNIQLEVARLSTEGNLEDKFPENTGHGHTIQLSANTLKON----- 803

Qy 131 VRLPKSLFRSLPGNRSVVRVLAATIL-DIGPGTLFKPRLGLG-----DGSGVLNNRLVG 183
Db 804 -----GRNGEIRVAFVLYNLLGPLYLSTENASMKLGTEAMSTNHSVIYNSPVIT 851

Qy 184 LSVQGMHVK--LAEPLIEVESH-QRPPNMTLCVFDVTKGT-TGDSSEGCs-TEVR 238
Db 852 AAINKFSNKVYLADPVVTVKHQSEENFNPCSEWSYKRTMTGYWSTQGCRLTTN 911

Qy 239 PEGTVCCCDHLTFALLLR--PTLDQSTVH--ILTRISQAGCGVSMIFLAFTIILYAFLR 294
Db 912 KTHITCSNHLTNFAVLMAHVEVKHSDAVHDLDDVITWVGILSLVCLLICIFTECF- 970

Qy 295 LSRRFRKSEDAKTHVALGSLFLNLAFLVNNGSGSGSDAACWARGAVFHYFLLCFT 354
Db 971 ----RGLQSDRNTHKNLCLSLFAVELLEL--IGINETDQPIACAVFAALLHFFFLAFT 1024

Qy 355 WNGLEAFHLYLLAVRVNTYFGH--YELKLSLVGWGLPALMVIGTGSAN--SYGLYTIIRD 410
Db 1025 WMELEGVQLXIMLVEVFESEHSRRKFEY---LVGYGMPALIVAVSAADVRSYG----- 1075

Qy 411 RENRTSLCLWFRGRTWYALYITVHGYPFLITP-----LFGMVVLALVWVKIFTLSR 462
Db 1076 -----TDKVCWRLDIT-----YFIWSFIGPATLIIMNLVIFLGIYALYKMF---H 1116

Qy 463 ATAV--KERK--KNRKK-VLTLLGLSLVGVTV--GLAIFTPLGLSTVYIFALFNSLQGV 515
Db 1117 HTAILKPESCLDNKISWVIGATIALCLLGLTWFAGFLMYNESTVIMAYLFTIFNSLQGM 1176

Qy 516 FICWFTIL 524
Db 1177 FIFIFHCVL 1185